

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

**University of Cape Town
Department of Physics**

**The evaluation of a research-based curriculum
for teaching measurement
in the first year physics laboratory**

Seshini Pillay

A dissertation submitted to the Faculty of Science at the University of Cape Town in fulfilment of the requirements for the degree of Master of Science in Physics.

March 2006

I declare that, except where acknowledged, this work is my own, completed with the guidance and advice of my supervisors.

Seshini Pillay

March 2006

University of Cape Town

Abstract

A new interactive introductory physics laboratory course, based on the ISO-recommended probabilistic interpretation of measurement and uncertainty has been developed. The course was piloted with first year physics students at the University of Cape Town in 2003. The present work evaluates this course. The sample cohort comprises approximately 150 GEPS students. These students are primarily from educationally disadvantaged backgrounds. Students' responses to diagnostic probes administered before and after participation in the course, are analysed in terms of the point and set paradigm framework. The level of consistent paradigm use by individual students across probes is investigated. The observed paradigm shifts are compared to those effected by a typical introductory physics laboratory course. The findings indicate a significant shift in students' understanding of measurement and uncertainty, across all aspects of measurement, to the set paradigm perspective. The success of the combination of interactive materials and the probabilistic approach to teaching measurement and uncertainty in the new course is discussed.

Acknowledgements

My thanks to Andy Buffler and Fred Lubben for their immeasurable contribution to this thesis and its completion.

This thesis is part of a project funded by the National Research Foundation.

University of Cape Town

Table of Contents

List of Tables	vii
List of Figures	ix
Chapter 1: Introduction	1
1.1 The role of the laboratory in physics curricula	1
1.2 Research into students' understanding of measurement and uncertainty	4
1.3 The point and set paradigm framework	11
1.4 The present research project	18
Chapter 2: The Intervention	20
2.1 The traditional physics laboratory course	20
2.1.1 Characteristics of the traditional physics laboratory course	20
2.1.2 Why the traditional course does not work	21
2.2 The new laboratory course	22
2.2.1 A probabilistic framework	22
2.2.2 Description of the new probabilistic laboratory course	26
Chapter 3: Method	34
3.1 Description of the sample	34
3.2 Research methods	35
3.3 Design of the questionnaire	36
3.3.1 Background of the questionnaire design	36
3.3.2 Experimental scenarios	37
3.3.3 Questionnaire probes	39
3.4 The questionnaires	41
3.5 Administration of probes	43

3.6 Analysis method	45
3.6.1 Construction of coding schemes	45
3.6.2 Reliability of code allocation	46
3.6.3 Individual probe analysis	46
3.6.4 Cross probe analysis	47
3.7 Interviews	47

Chapter 4: Findings - Individual Probe Analysis **48**

4.1 Data collection probes: Students' ideas about repeating measurements	48
4.1.1 The RD and RDA probes	48
4.2 Data processing probes: Students' ideas about single measurements	51
4.2.1 The QD1 and PR2 probes	51
4.2.2 The QD2 and PR1 probes	57
4.2.3 The QD3 and ED1 probes	62
4.3 Data processing probes: Students' ideas about the average of a set of measurements	
4.3.1 The UR probes	66
4.3.2 The UA probes	68
4.4 Data comparison probes: Students' ideas about comparing data sets and averages	
4.4.1 The SMDS and DMSS probes	76
4.5 Understanding uncertainty probes: Students' ideas about uncertainty in measurement	
4.5.1 The NU1 and PX1 probes	79
4.6 Nature of measurement probes: Students' ideas about the nature of measurement	87

Chapter 5: Findings - Cross Probe Analysis **93**

5.1 Understanding of measurement in data processing	93
5.1.1 Single measurements	94
5.1.2 Multiple measurements	95
5.2 Understanding of measurement in data comparison	97
5.3 Understanding measurement uncertainty	100
5.4 Cross-probe analysis for all probes	102

Chapter 6: Discussion	104
6.1 Critical reflections on the research methods	104
6.2 Answers to the research questions	106
6.2.1 Students' perceptions of measurement and uncertainty	106
6.2.2 Effectiveness of the new laboratory course	108
6.3 Findings related to previous studies	109
6.3.1 Comparison of the effectiveness of the old and new laboratory courses	109
6.4 Why the new course is better	113
6.5 Implications for teaching	115
6.6 Further work	116
References	117
Appendices	125
Appendix I: The complete pre-intervention measurement questionnaire	125
Appendix II: The coding schemes for the pre-intervention probes	141
Appendix III: The complete post-intervention measurement questionnaire	160
Appendix IV: The coding schemes for the post-intervention probes	171

List of Tables

Table 1.1: Model of progression of ideas concerning experimental data (Adapted from Lubben and Millar, 1996)	11
Table 1.2: The point and set paradigms (Buffler et al., 2003)	12
Table 1.3: Actions and reasoning associated with the point and set paradigms (Buffler et al., 2001)	15
Table 2.1: Outline of the content of the interactive student workbook	27
Table 3.1: List of probes used in this study	43
Table 4.1: Students' paradigm use in their responses to the to the pre-intervention RD and RDA probes ($n = 100$)	50
Table 4.2: Students' paradigm use in their pre- and post- intervention responses to probes dealing with a single reading on an analogue scale (reference point appears directly below calibration line) ($n = 76$)	53
Table 4.3: Students' paradigm use in their pre- and post- intervention responses to probes dealing with a single reading on an analogue scale (reference point appears between calibration lines) ($n = 76$)	59
Table 4.4: Students' paradigm use in their pre- and post- intervention responses to probes dealing with a single reading on a digital Scale ($n = 76$)	63
Table 4.5: Students' paradigm use in their pre- and post- intervention responses to the UR probe ($n = 76$)	67
Table 4.6: Students' paradigm use in their pre- and post- intervention responses to the UA1 Probe ($n = 76$)	69

Table 4.7: Students' paradigm use in their pre- and post- intervention responses to the UA2 probe ($n = 76$)	73
Table 4.8: Students' paradigm use in their pre- and post- intervention responses to the SMDS probe ($n = 76$)	77
Table 4.9: Students' paradigm use in their pre- and post- intervention responses to the DMSS probe ($n = 76$)	79
Table 4.10: Students' paradigm use in their responses to the pre-intervention NU1 probe ($n = 100$)	80
Table 4.11: Students' paradigm use in their responses to the post-intervention PX1 probe ($n = 117$)	84
Table 4.12: Students' paradigm use in their responses to the pre-intervention AE1 probe ($n = 100$)	88
Table 5.1: Pre- and post- instruction paradigm distributions among students when processing a single reading ($n = 76$)	94
Table 5.2: Pre- and post- instruction paradigm distributions among students when processing multiple readings ($n = 76$)	96
Table 5.3: Pre- and post- instruction paradigm distributions among students when comparing data sets ($n = 76$)	98
Table 5.4: Pre- and post- instruction paradigm distributions among students when considering uncertainty in measurement ($n = 76$)	100
Table 5.5: Pre- and post- instruction paradigm distributions among students across all probes ($n = 76$)	102
Table 6.1: Comparison of paradigms used by students after participation in the original and new GEPS laboratory courses	113

List of Figures

Figure 1.1: The goal of instruction in relation to the point and set paradigms (Buffler <i>et al.</i> , 2001)	17
Figure 2.1 : A model for determining the result of a measurement (Allie <i>et al.</i> , 2004)	24
Figure 2.2: An example of an ‘authentic’ problem based practical exercise	27
Figure 2.3: An example page from the student workbook (Buffler <i>et al.</i> , 2002)	28
Figure 2.4: Another example page from the student workbook (Buffler <i>et al.</i> , 2002)	29
Figure 3.1: Description of experimental context as it appears on pre-intervention questionnaire covers	38
Figure 3.2: Description of experimental context as it appears on post-intervention questionnaire covers	39
Figure 3.3: The RD (Repeating Distance) probe	40
Figure 5.1: Comparison of paradigm distributions for probes relating to a single reading before and after instruction	95
Figure 5.2: Comparison of paradigm distributions for probes relating to the average of a data set before and after instruction	97
Figure 5.3: Comparison of paradigm distributions for probes relating to comparing data sets before and after instruction	99
Figure 5.4: Comparison of paradigm distributions for probes related to uncertainty in measurement before and after instruction	101
Figure 5.5: Comparison of paradigm distributions over all probes before and after instruction	103

Figure 6.1: Students' paradigm use when processing data after participation in the old and new laboratory courses	110
Figure 6.2: Students' paradigm use when comparing data after participation in the old and new laboratory courses	111
Figure 6.3: Students' paradigm use across all measurement areas after participation in the old and new laboratory courses	112

University of Cape Town

Chapter 1: Introduction

A collaborative research programme between the science education research groups from the Department of Physics at the University of Cape Town (UCT) in South Africa and the Department of Educational Studies at the University of York (UOY) in the United Kingdom, was initiated in 1995 with the primary intention of investigating physics students' understanding of measurement, and to thus determine a theoretical basis for the development and implementation of a new introductory physics laboratory curriculum. This curriculum, while evolving to assimilate changes in international conventions for reporting scientific measurement, would improve students' proficiency in procedures of scientific measurement and data analysis, and cultivate their understanding of measurement and uncertainty.

The laboratory course evaluated in this study is a product of the UCT - UOY collaboration.

1.1 The role of the laboratory in physics curricula

Notwithstanding a few dissenters, there has long been a general consensus among science researchers and educators that practical work is an essential component of teaching science. It is thus that laboratory courses form part of most undergraduate physics curricula. However, since the introduction of the laboratory component to science teaching in the mid 1800's (Gee and Clackson, 1992), there has been little consensus about the purposes of this practical element (White, 1996). The aims of laboratory courses are often manifold and confusingly combined. They include demonstrating physical principles introduced in lectures, providing "hands-on" opportunities to familiarize students with experimental procedures and apparatus, and introducing measurement and data analysis techniques (Gott and Duggan, 1996). Clearly defined overall themes are rarely evident in laboratory curricula.

In order to isolate the main objectives of a laboratory course, it is essential that the role of the laboratory course in teaching and learning science is understood.

Millar (1996) suggests that acquiring scientific knowledge, or learning science, is a multi-dimensional enterprise. It begins with consolidating and expanding *declarative* knowledge (Black, 1993), or understanding of the accepted concepts, phenomena and laws of science, then moves on to developing *procedural* knowledge (Millar *et al.*, 1994), which includes collecting data through observations and measurements, analyzing and interpreting the data collected, and comparing the conclusions drawn from the data analysis to an existing knowledge base. The final step to acquiring scientific knowledge is establishing relevant links between science and society to provide a meaningful context for the results found (Vygotsky, 1978).

It is important to note that there is no single agreed definition for the term “procedural knowledge”. Its use in the context of experimental work in science, as detailed earlier, is distinct from its description of the ability of students to apply algorithmic procedures when solving written problems, as employed by Larkin and Reif (1989) and Chi *et al.* (1981).

In their study of the effectiveness of school laboratory curricula in Britain in 1994, Millar *et al* identified three areas of procedural understanding when undertaking a scientific investigation: (i) students’ perceptions of the purpose of the investigation which influences their interpretation of what the experiment involves; (ii) students’ skill in the use of experimental apparatus and in applying experimental techniques which determines the quality of the data acquired; and (iii) students’ understanding of “concepts of evidence” (Gott and Duggan, 1996) which allows them to judge the quality (both the reliability and the validity) of investigative process based on the ideas they hold about the evaluation criteria.

It is further suggested by Gott and Duggan (1996) that the adopted methods of data collection, data presentation and data interpretation are all impacted upon by the students’ perceptions of the validity and reliability of experimental procedures. The new knowledge obtained from the experimental investigation must then be passed to the realm of shared scientific knowledge through careful consideration of the quality of the consolidated result, and its unambiguous communication (McGinn and Roth, 1999).

Most undergraduate laboratory courses comprise a series of highly structured verification experiments designed to increase students' declarative knowledge, i.e. students' understanding of the concepts, laws and models introduced in lectures (Meester and Maskill, 1995; Laws, 1996; Tiberghien *et al.*, 2001). However, various studies have questioned the effectiveness of this type of hands-on experiment for illustrating theory and phenomena (Roth *et al.*, 1997; Kirchner and Huisman, 1998), considering that they are essentially mock experiments with pre-determined outcomes. These experiments distort students' perceptions of the scientific approach to enquiry, and promote the idea that a "perfect" experiment will yield a "perfect" result (Hodson, 1998). In spite of this, teachers continue to use verification experiments since they are easy to set up for small and large groups of students, and the predictable results ensure that they are also easy to assess (Montes and Rockley, 2002).

An alternative laboratory course structure proposed by Etkina *et al.* (2002) suggests that the laboratory exercises should be built around three types of experiments, viz. observational experiments, testing experiments and application experiments. The observational experiments focus on demonstrating new phenomena for which students offer possible explanations; the testing experiments verify predictions based on previously devised explanations of the same phenomena; and the application experiments use the explanation of one phenomenon to predict another. This structure, called the "process approach", while offering new laboratory experiences, still emphasizes the development of concepts, laws and models.

Osborne (1996), rather than presenting another differently structured laboratory course with the same emphasis, propounds that it is the purpose of hands-on laboratory exercises which should be altered to focus more strongly on developing a scientific approach to enquiry. This approach would be characterized by the students' ability to: (i) prepare and perform a standard experimental procedure using familiar apparatus; (ii) plan an experimental investigation to address a given task; (iii) collect, process and compare data; (iv) relate evidence to theory, (or use data to support a conclusion); and (v) clearly communicate the results of the experimental investigation (Millar *et al.*, 1999).

Leach (1999), in his investigation of students' understanding of the co-ordination of theory and evidence in science, was able to refute the claims of Kuhn *et al.* (1998), that younger students (those in the 9 - 16 year age group) are unable to evaluate knowledge claims, or theories, in terms of experimental evidence. His findings confirm that these students are capable of coordinating theory and evidence, but lack knowledge of the rules of theory evaluation in science. He advocates the explicit teaching of how to generate predictions and evaluate observations in terms of stated theories, and the social and empirical validation of scientific knowledge.

Although the body of literature describing students' procedural knowledge (Roth and Roychoudury, 1993; Germann and Aram, 1996) is significantly smaller than that chronicling students' declarative knowledge (summarized by Pfundt and Duit, 1994), the implication of the research summarised here is that procedural knowledge is not a collection of skills to be practised, but is rather a distinct domain of knowledge to be learned. It follows then, that this domain of knowledge must be taught. This, in turn, necessitates the restructuring of laboratory courses to facilitate the explicit teaching of the fundamentals of collecting, processing, comparing, interpreting and presenting data. In order to develop a theoretical framework around which such a research-based laboratory course can be constructed, it is imperative to gain some insight to students' understanding of measurement and its associated uncertainties.

1.2 Research into students' understanding of measurement and uncertainty

A study of the ideas held by French students in their first year of university, about measurement, measurement errors and the statistical analysis of measurement errors, upon completion of a theoretical course on data analysis was undertaken by Séré *et al.* (1993). The data analysis course focused on imparting conceptual knowledge of ideas associated with measurement, i.e. 'true' value, precision, accuracy, dispersion, error, uncertainty, and on introducing mathematical tools including mean values, standard deviations and Gaussian distributions. The study revealed that while able to adequately apply algorithms for calculations of means, standard deviations and confidence intervals, most students still

lacked an understanding of the statistical procedures followed in completing a practical investigation. Measurements were repeated only to confirm or validate the initial measurement; repeated measurements were treated as a series of single observations each with its own confidence interval; making many measurements was considered desirable for a “better” result, with no clear notion of the nature of “better”; the first or recurring measurement result was reported as the final answer; there was no comparison of, or comment on, results; and students’ displayed difficulty in distinguishing between the concepts of precision and accuracy.

This confusion about the terminology and underlying concepts of precision, accuracy, systematic error and random error is mirrored in Garrett *et al.*’s (2000) study of first year chemistry students in the United Kingdom. Tomlinson *et al.* (2001) suggested the use of a well-defined set of key words to alleviate confusion in students’ laboratory reports. However, Thomson (1997) had earlier pointed out that terminology is inconsistently used, even in physics publications.

In an exploratory study involving a small group of students at the University of Leeds, Ryder and Clarke (2001) found that students’ understanding of the terms “systematic error”, “random error”, “precision” and “accuracy” is significantly enhanced by explicit teaching about sources of error. Prior to participation in a ten week course on sources of error, students displayed confusion about the terms mentioned, but were able to offer clear definitions for them after instruction. The findings also highlighted the fact that students blindly employed computational methods when analyzing data, without consideration of the origin of the data or the implications for its analysis.

In their study of the reasoning followed, and the difficulties encountered, by 14 - 17 year old secondary school students in France during a measurement activity, Coelho and Séré (1998) broached the concept of a “true value” through interviews covering data collection, processing and interpretation. Students’ conceptions were classified as either advantages, or obstacles, to relevant measuring processes, depending on the nature of the teaching and learning activities with which students engaged. Students’ responses and actions were influenced by the level of their beliefs in a “true value”. Most students endeavoured to find this “true value”, and expressed discontent with varying measurement results. This belief

in one true value was viewed by the authors as a “double-edged sword” simultaneously taking on the role of precursor to a deeper understanding of measurement and posing an obstacle to the same. On the one hand, the quest for the true value encourages students to improve their experimental techniques and to repeat measurements in the attempt to eliminate variability. On the other hand, the desire to eliminate variability in measurement promotes the idea that it is possible to attain a “perfect” measurement with no associated uncertainty given the ideal method and apparatus. Unfortunately, the nature of instruction in traditional laboratory courses makes them more conducive to the latter outcome.

Fairbrother and Hackling (1997) suggest that one of the problems with traditional laboratory is the closed nature of many practical exercises, which they claim is rooted in the epistemological view of science as a body of knowledge to be catalogued. These exercises amplify students’ expectation of a ‘right’ answer to an experimental problem. Variations in measurements and deviations from the expected answer lead students to believe that they have made mistakes or ‘errors’ in their approach to the experiment, and where the concept of uncertainty is not clearly understood, students deem it possible to completely eliminate these ‘errors’. The implication is that most traditional laboratory courses are designed to fortify rather questionable epistemological tenets, with no consideration of students’ preconceptions about measurement in science. This serves only to impede the effective learning and teaching of scientific measurement.

The perception of science as an inventory of facts to be learned hinders students’ understanding and acceptance of new concepts. While easily able to reason through and make sense of everyday phenomena, they are resistant to applying similar reasoning to the relatively unfamiliar realm of science. Reif and Larkin (1991) point out that in addition to students’ inconsistent reasoning between everyday and scientific contexts, their distorted views of the nature of scientific knowledge are enhanced by the fact that science taught at schools is different from both everyday scenarios and real scientific contexts.

In Hammer’s (1994) investigation of the epistemological beliefs of a cohort of undergraduate physics students, he identified three categories of understanding of the nature of knowledge and learning: (i) the structure of physics - students perceived physics as either a collection of isolated information or as a coherent framework; (ii) the content of

physics - students held the belief that physics contains a quantity of facts and formulae which need to be remembered, or concepts to be assimilated and applied; and (iii) the learning of physics - students viewed learning physics as either the receiving and processing of information, or as the developing of their own understanding of information. Hammer concluded that students' beliefs about the nature of knowledge and learning affected their success in learning physics. Elby (2000) concurred, noting that traditional laboratory courses have little success in changing students' epistemological beliefs, which affect their metacognitive practices and their study habits. He proposed an epistemology focused course designed to apprise students of the validity of the application of common sense, not only in everyday contexts, but also in scientific thinking. Hammer and Elby (2003) reviewed literature evidencing the importance of this epistemological component to successful learning in introductory physics courses, and observed that high school students form robust, yet counter productive, epistemological views about science. In particular, students' view of science as a catalogue of facts prompts them to the belief that an unexpected measurement result, or failure to verify a fact, is a consequence of experimental or 'human' error. This is supported by the reports of Ryder and Leach (2000) and Leach *et al.* (2000) on a study, involving close to 800 senior secondary and university students in five European countries, of students' data interpretation skills. It was found that students use many forms of epistemological reasoning in their interpretation of data, while ignoring theoretical models. They suggest that it is essential to consider the multiple forms of epistemological reasoning when designing curricula.

The results of Séré *et al.*'s (2001) investigation into the nature of understanding of measurement held by senior secondary and first year university students, reinforced Leach *et al.*'s (2000) finding that students' decisions when dealing with data are informed by more than just one form of epistemological reasoning. The investigation took the form of a diagnostic questionnaire administered to approximately 400 students in France and Spain. The questionnaire explored students' reasoning when handling sets of experimental data, with the aim of gauging the extent of the influence of their epistemological beliefs on their claims about the data sets. The authors concluded that students' decisions at the different stages of a laboratory task - determining what constitutes a reliable measurement, choosing a suitable measurement technique, processing measurements and interpreting processed measurements - were rooted in different epistemologies depending on the context.

Another study of undergraduate students' handling of experimental measurement was undertaken by Evangelinos *et al.* (1998). The main focus of this study was on how students deal with single readings. The results indicated that measurements were repeated only for verification, and the necessity for repetition of measurements was determined by the measuring apparatus. When using a digital measuring device, which students consider a high precision instrument, a single reading was deemed sufficient for finding the true value of the quantity being measured (measurand). This view persisted even after instruction. Digital readings were seen as exact, and the notion of precision was associated with the number of digits on the display. The students' deep-seated views of exactness and precision acted as barriers to their acceptance of uncertainty as an intrinsic property of scientific measurement.

Evangelinos *et al.* (2002) subsequently carried out an intervention study, with a sample of first year university students in Greece, using the probabilistic approach to measurement. The students were classified, according to their perceptions of the relationship between the measurand (variable being measured), or theory, and the measurement (datum), or evidence, as "exact", "approximate" or "interval" reasoners. "Exact" reasoners subscribed to the notion of a 'good' single measurement representing an exact value; "approximate" reasoners, upon realization that an ideal measurement is unobtainable, considered it most appropriate to represent a single measurement as an approximate value; and "interval" reasoners deemed intervals the best representation of measurements only in the event of really 'bad' measurements. The results of the study suggested that the intervention increased students' understanding of the fundamental difference between an exact quantity and an uncertain one, and facilitated their learning of how to apply the concepts of uncertainty and probability to single measurements.

The comparison of two data sets was the focus of an interview survey of American undergraduate students, carried out by Masnick and Morris (2002). Their aim was to determine what influence the characteristics of the data sets - the number of data points in the set, the frequency of overlapping data points and the range of the data points relative to their mean - had on how students compared them. Students were asked what conclusions they were able to draw from the given data sets, how they reached these conclusions, and

how certain of them they were. They were then required to predict the next data point for each set, and to comment on their level of certainty of the difference between the predicted values. It was found that students' responses were greatly influenced by the sample size and by the number of overlapping data points. Students were significantly more confident of their conclusions and predictions when presented with larger samples, and when there were fewer overlapping data points students expressed a greater certainty of the difference between the data sets. While most students also considered factors such as the means of the data sets when forming their conclusions, only a very small number noted the possible influence of variability or outliers within the data sets and the experimental method and apparatus in their reasoning.

The Scientific Community Laboratory (SCL), an intervention for teaching measurement and uncertainty to physics undergraduates in the USA, was evaluated by Lippmann (2003). The SCL approach encourages students to draw on their everyday deductive and reasoning skills when making decisions related to data collection and interpretation in the laboratory. This method demonstrated success in facilitating students' understanding of the relevance of intervals in the comparison of data sets.

Deardoff (2001) recently undertook a study of introductory physics students' conceptions of measurement uncertainty and error analysis. Data was collected through written surveys and interviews from sample cohorts from two universities in the USA, and one in Japan. The findings indicated that students, across the institutions, lacked an understanding of the significance of uncertainty in measurement. Uncertainty estimates were ignored when assessing agreement between results; students tended to avoid making quantitative statements of uncertainties associated with a measurement; and they exhibited difficulty in identifying and quantifying possible sources of error. Encouragingly, it was found that the quality of students' understanding was aligned with the amount of instruction they had received.

Abbott's (2003) investigation into the change in students' understanding of measurement and uncertainty after participation in a semester-long introductory physics laboratory course revealed that explicit instruction on measurement practice and uncertainty calculation leads to improvement in student understanding. His study focused on 500 North

American students, only half of whom received instruction in measurement and uncertainty. The results of several probes into various aspects of measurement, including the meaning of the spread of data, the significance of repeating measurements and reading analogue scales, suggest while the traditional teaching methods do contribute to improving students' understanding, many students leave the introductory physics laboratory without appreciation for, or coherent understanding of, the concept of measurement uncertainty. This is in agreement with the findings of Deardoff (2001).

The introduction of a new national curriculum in England and Wales in the 1990's prompted the initiation of the Procedural and Conceptual Knowledge in Science (PACKS) Project (Millar *et al.*, 1994), which gauged the effectiveness of school laboratory curricula. The PACKS project observed the actions of children in the 9 - 14 age group when presented with open-ended investigative tasks, and solicited their reasoning through diagnostic questions or 'probes', to develop a model linking students' actions to their understanding of measurement. It was found that students base their decisions in the tasks not only on their understanding of the relevant science concepts, but also on their perceptions of the purposes of the investigations and their ideas about the quality and validity of empirical data.

The second phase of the PACKS project (Lubben and Millar, 1996) focused on English secondary school and pre-university students, and delved into students' understanding of the validity and reliability of measurements. A result of this study was a suggested model for the progression of students' ideas about measurement (Table 1.1).

The model was based on students' responses to a series of pencil-and-paper exercises probing students' measurement actions in many different experimental scenarios. The authors emphasize that the progression through the levels is not a reflection of students' progressive learning paths. The model does, however, provide a framework for classifying measurement actions in terms of the underlying measurement concepts.

Table 1.1: *Model of progression of ideas concerning experimental data (adapted from Lubben and Millar (1996)).*

Level	Students' view of the process of measuring
A	Measure once and this is the right value.
B	Unless you get a value different from what you expect, a measurement is correct.
C	Make a few trial measurements for practice, and then take the measurement you want.
D	Repeat measurements till you get a recurring value. This is the correct measurement.
E	You need to take a mean of different measurements. Slightly vary the conditions to avoid getting the same results.
F	Take a mean of several measurements to take care of variation due to inaccurate measuring. Quality of the result can be judged only by authority source.
G	Take a mean of several measurements. The spread of all the measurements indicates the quality of the result.
H	The consistency of the set of measurements can be judged and anomalous measurements need to be rejected before taking a mean.

1.3 The point and set paradigm framework

The Lubben-Millar model presented in Table 1.1 is a descriptive schema which, while useful for identifying the different levels of sophistication attained by students, fails to provide explanations for students' actions and responses in experimental exercises.

In a research study undertaken by Allie *et al.* (1998), students' responses to diagnostic questions probing various aspects of measurement revealed links between their routes of reasoning and their perceptions of measurement as either single 'true' values, or as a spread of values. This led to the classification of students as "point reasoners" or "set reasoners", and to the development and definitions of the point and set paradigms, as summarized in Table 1.2.

Table 1.2: *The point and set paradigms (Buffler et al., 2003).*

Point Paradigm	Set Paradigm
The measurement process allows you to determine the true value of the measurand.	The measurement process provides incomplete information about the measurand.
“Errors” associated with the measurement process may be reduced to zero.	All measurements are subject to uncertainties that cannot be reduced to zero.
A single reading has the potential of being the true value.	All available data are used to construct distributions from which the best approximation of the measurand and an interval of uncertainty are derived.

The point paradigm is characterized by the notion that each measurement could potentially yield the correct or “true” value of the quantity being measured (measurand), or an incorrect value. Consequently, each measurement is considered independent of every other measurement, except where one value obtained is used as the basis for the confirmation or rejection of another. Adherents to the point paradigm subscribe to the belief that a single carefully performed measurement is sufficient to establish the one true value of the measurand. Deviation from an expected result is attributed to mistakes in the experimental procedure. Given an ensemble of readings with dispersion, a representation of the measurement is selected, not by consideration of the ensemble as a whole, but by inspection of the individual data points. For example, in the case of one data set the highest value, the lowest value or the recurring value is chosen as the final result, and in the case of more than one data set, agreement, lack thereof, is determined by a one-to-one comparison of data values.

The set paradigm, on the other hand, is characterized by the notion that each reading is an approximation of the measurand, and that in principle knowledge about the measurand can never be complete or perfect. All available data are used to construct distributions from which the best approximation or estimate of the measurand and an interval of uncertainty are derived. In most cases, the best approximation of the measurand will be the reading itself (for a single reading), or the average of readings (for an ensemble with dispersion).

This, together with the combined uncertainties, is used to determine confidence intervals and to facilitate the comparison of different data sets. It can be said that the point paradigm is a 'local realistic way of viewing data', while the set paradigm 'uses theory to mediate between the data and the measurand' (Campbell *et al.*, 2005).

To test the extent to which the point and set framework was useful for interpreting students' ideas about measurement and uncertainty, Lubben *et al.* (2001) undertook a study to classify students' ideas prior to instruction in terms of the point and set paradigms. The sample cohort consisted of both mainstream and GEPS UCT students entering their first year of university study. (Students who meet the minimum entrance requirements enter directly into the standard three-year degree programme, also referred to as the mainstream or "direct entry" programme. Students from educationally disadvantaged backgrounds who fail to meet the minimum entrance requirements enter a four-year degree programme, the General Entry to Programmes in Science (GEPS), or the "bridging" programme.) Probes identical to those employed in the Allie *et al.* (1998) study were used to explore students' ideas about data collection, data processing and data comparison.

The results of the study showed that if the students' responses were classified according to the point and set paradigms, there was a good correlation between the reasoning used across the data collection probes and that adopted for data processing decisions. This trend was reinforced by the classification of students' who calculated the mean as a rote-learned routine as point reasoners.

It was found that the reasons for repeating measurements, the ways of dealing with a collection of repeated measurements and the fitting of a straight line to a set of plotted points, were all rooted in a common construct. For example, students who subscribed to point-based reasoning typically did not repeat measurements except in attempts to improve their experimental skills or to find recurring values; they often chose a recurring value in a data set as a representation on the set; and when required to fit a straight line to a series of plotted points, they typically opted to connect the points with a series of straight line segments, or to draw a straight line through as many points as possible. In contrast, set reasoners tended to represent a data set by a calculated mean and, in some cases, an

estimation of uncertainty; and tried to taken into account all data points when fitting straight lines, usually by ensuring the same number of points above and below the line.

The alternate usage of point and set reasoning by the same students for different probes within the stages of data collection and data processing occurred frequently. It was noted, however, that this usage was not random but rather related to the procedural context. For example, many consistent point reasoners adopted set reasoning when dealing with measurements of time. While content to use a recurring value for distance, these students, apparently prompted by the variability of time measurements due to the operation of the stopwatch, stated that repeated measurements were required for the calculation of the average time.

Song and Black (1992) reported that practical performance depended on the conceptual demand of the science context and the laboratory-versus-everyday context of the investigative task. The finding that students' use of either the point paradigm or the set paradigm is dependent on the procedural context, led Lubben *et al.* (2001) to go even further to the conclusion that measurement decisions also depend on the measurement context of the task.

However, even among the 25% of students who displayed consistently set-based reasoning across data collection and data processing probes, a fully internalised understanding was still lacking. The data comparison probes required the students to realise that both the mean of a series of individual data points and degree of dispersion are essential characteristics of the data set. The findings showed that while a good proportion of the students recognised the inherent spread in the data, they used only the mean to represent the set. The implication is that at this high level of measurement demand, set reasoning is only maintained at a low cognitive level, *i.e.* recognition.

Germann *et al.* (1996) noted that students' actions were not always consistent with their stated reasoning about measurement. The Lubben *et al.* (2001) study found similar examples of students who appeared to use both point and set reasoning in a fragmented way. For example, some students stated that they repeat measurements in order to take an average but, when offered the option, they selected the recurring value to represent the

series of measurements; others used the term “averaging” and correctly described the process of calculating the average, but then chose a reading closest to the calculated average to represent the measurement. This contradiction between reasoning and action was also apparent in the responses to the straight line graph probe where some students described an appropriate procedure to fit a line to the data, but then drew a line segment through as many data points as possible.

These findings support the suggestion put forth by Buffler *et al.* (2001) that students need to acquire proficiency in both the tools and procedures of data analysis (actions) and the understanding of the nature of scientific measurement (reasoning). Table 1.3 summarizes the actions and reasoning associated with the point and set paradigms.

Table 1.3: *Actions and reasoning associated with the point and set paradigms (Buffler et al., 2001).*

Point paradigm		
Measurement phase	Action	Reasoning
Data collection	No repeating of measurements is necessary, or repeat to find recurring value, or repeat for practice.	A measurement leads to a single, “point-like” value rather than contributing to an interval. Only one good measurement is required.
Data processing (Calculation)	A single (best) measurement, e.g. the recurring value, is selected to represent the true value.	Each single measurement is independent of all others and can in principle be the true value.
Data processing (Straight line graph)	All points joined by multiple line segments or a single line drawn through selected data points.	The trend of the data is best represented by selecting particular data points which describe the desired trend.
Data set comparison	A value-by-value comparison of the two sets, or comparison based on the “closeness” of the means (if given).	No basis for the need to repeat measurements therefore comparisons made on the basis of the closeness of individual points.

Set paradigm		
Measurement phase	Action	Reasoning
Data collection	Repeating of measurements of the same quantity is necessary as a consequence of the inherent spread in data.	Each measurement is only an approximation to the true value and that the deviation from the true value is random. A large number of measurements are required to form a distribution that will cluster around some particular value.
Data processing (Calculation)	A set of measurements is represented by theoretical constructs, e.g. the mean and standard deviation.	The best information regarding the true value is obtained by combining the measurements using theoretical constructs in order to characterise the set as a whole.
Data processing (Straight line graph)	All the measurements taken into account by a least squares straight line fit to all the data.	The best graphical representation of series of measurements is obtained by modelling the trend of the data.
Data set quality	For the same number of measurements, the better measurement is chosen to be the one associated with the smallest standard deviation.	The standard deviation is related to the precision of the measurement.
Data set comparison	The agreement of two measurements is related to the degree of the overlap of their intervals.	The mean and standard deviation define a confidence interval which is related to both the best estimate and the reliability of the measurement.

This study also highlighted the fact that students' actions and reasoning can be drawn from either the point or the set paradigm on an ad hoc basis depending on the procedural context. Buffler *et al.*'s (2001) illustration, reproduced in Figure 1.1, shows the four main categories into which students can be classified according to their actions and reasoning.

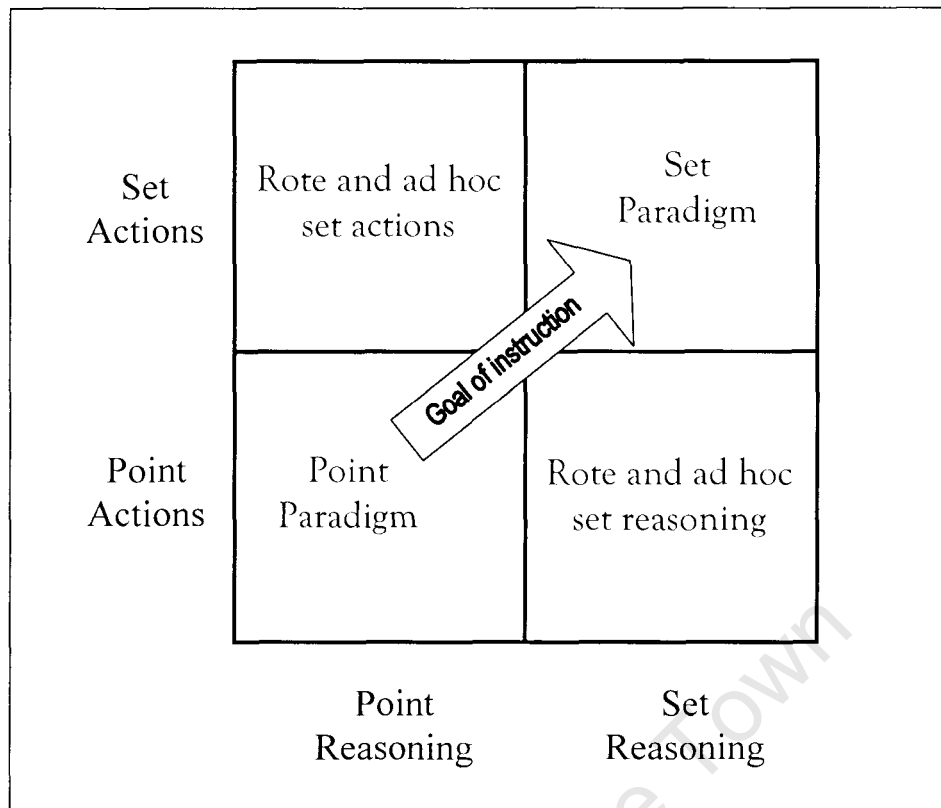


Figure 1.1: *The goal of instruction in relation to the point and set paradigms (Baffler et al., 2001).*

The bottom left-hand quarter houses students whose reasoning and actions are both firmly rooted in the point paradigm, while students who consistently reason and act according to the set paradigm reside in the top right-hand quarter. These opposite corners represent the pure point and set cases as described in Table 1.3.

Students who employ the tools and actions associated with the set paradigm by rote (i.e. they are able to complete set-based data analysis procedures, while retaining a theoretical understanding rooted in the point paradigm) fall into the top left-hand quadrant. The bottom right-hand quadrant is reserved for those students who have a set paradigm-based view of measurement but have yet to master the tools and procedures of data analysis.

It follows that the broad purpose of laboratory instruction is to facilitate a shift in students' paradigm use for understanding scientific measurement to the top right-hand corner.

1.4 The present research project

An introductory physics laboratory course developed for GEPS students at UCT (Allie and Buffler, 1998) was evaluated, in terms of students' use of paradigms, by Buffler *et al.* (2001). The results of the evaluation indicate that the course was not successful in effecting any significant shift in students' paradigm use from the point to the set paradigm when dealing with scientific measurement. Although the course was able to achieve its aim of teaching students' the formal procedures of data analysis, it failed to develop in students, an appreciation of the links between the nature of measurement and data processing methods. This shortcoming of the course was evidenced by students' ability to perform routine procedures based on the set paradigm, while reverting to the point paradigm when required to reason through a given task.

The indication that the established GEPS laboratory course was largely ineffectual in achieving the "goal of instruction" illustrated in Figure 1.1, together with an awareness of the discrepancies between the conventions for reporting on experimental measurements in physics research and the content of undergraduate laboratory courses motivated the UCT - UOY collaborators to design and implement a new laboratory course with an emphasis on measurement and associated uncertainties. (The framework on which the new course is based, and a detailed description of the course, follow in Chapter 2.)

The evaluation of this new course is the foundation of the work presented in this discourse. The results of the evaluation will expose the deficiencies of the course in terms of its goals, and also highlight its successes. This, in turn, will facilitate the modification, amendment and extension of the course to address its shortcomings and, in particular, to accommodate students' pre-conceptions of measurement and to eliminate their misconceptions. It will also influence the teaching and tutoring practices in the physics laboratory.

The purpose of this study is to evaluate the new laboratory course by answering the following questions:

- What ideas do first year physics students have about measurement and uncertainty, and how do these ideas differ before and after participation in the new laboratory course based on the probabilistic framework for measurement and uncertainty?
- How effective is the new laboratory course in changing students' understanding of measurement and uncertainty in terms of the point and set framework?

University of Cape Town

Chapter 2: The Intervention

2.1 The traditional physics laboratory course

2.1.1 Characteristics of the traditional laboratory course

A primary goal of a typical first year physics laboratory course is to develop students' understanding of the measurement process and their skills in the use of measuring instruments and data analysis tools, through 'hands-on' experience in various experimental situations, demonstrations of concepts and phenomena introduced in lectures, and instruction in the basics of scientific measurement. The experiments are usually presented as recipe-type laboratory exercises which students are required to work through systematically. This well-known laboratory practical format has been shown to be ineffectual in achieving the afore-mentioned aims (Allie *et al.*, 1997).

The common premise on which most laboratory curricula are based is that the mechanical application of standard procedures will result in students acquiring an understanding of the nature of measurement and experimentation. This assumption has, however, been challenged by studies (Sere *et al.*, 1993, Giordano, 1999) which highlight the fact that although students having completed a laboratory course generally demonstrate an adequate grasp of mechanistic data analysis techniques, they often display a remarkable lack of appreciation of the nature of scientific measurement. In particular, the concept of uncertainty is not well understood by many of the students. For example, in the evaluation of the original UCT GEPS physics laboratory course (Buffler *et al.*, 2001) it was found that while the majority of students were easily able to calculate the mean of a set of readings, few were cognizant of the fact that the mean has little significance without some indication of the corresponding uncertainty.

Evaluations of other similarly based laboratory courses - a chemistry laboratory course for first year students in a bridging programme (Rollnick *et al.*, 2002) and a physics laboratory course for mainstream first year students (Volkwyn *et al.*, 2004) - yielded similar results.

On entering university, most students view scientific measurement from the point paradigm perspective, subscribing to the notion that a ‘correct’ measurement is one that has no associated uncertainty (Lubben *et al.*, 2001). Studies (Buffler *et al.*, 2001) reveal that even after participation in a laboratory course carefully structured to shift students’ reasoning from the point paradigm to the set paradigm, the majority of students retain the idea of an ‘exact’ or ‘point-like’ value.

The overwhelming implication is that traditional laboratory courses, while successful in imparting the formal data analysis techniques, fail to develop students’ understanding of the relationship between these analysis methods and the nature of measurement.

2.1.2 Why the traditional course does not work

Traditional laboratory courses are often structured as series of experimental tasks demonstrating concepts introduced in lectures, often reproducing well-known results. Experimental methods and data analysis are not taught explicitly, but rather addressed only when questions arise during experiments. When broached, the subject of measurement is presented as a “combination of rigorous mathematical computations and vague rules of thumb” (Buffler *et al.*, 2004).

Traditional data analysis in introductory physics courses is typically based on the frequentist approach to statistical analysis. This method assumes that there is a true value for a measurand. Measurements are thus made to determine the true value, and each measurement has some associated random scatter. This scatter is usually represented by a Gaussian distribution, the mean value of which tends to the true value as the number of measurements increases. The frequentist approach asserts that the true value is a fixed, unknown constant with no associated uncertainty. It is the data which are uncertain due to the inherent randomness of the measurement process. Rigorous mathematical models are employed for dealing with random errors, while systematic errors are reduced to unknown constants to be determined by examining the experimental setup.

The success of this statistical approach is dependent on large sample sizes. However, the number of readings taken in most first year laboratory experiments is typically limited to five or six because of time constraints. Furthermore, frequentist statistics does not offer a logical way to deal with a single measurement.

The inconsistencies in dealing with different sources of error, and the *use* of the term ‘error’ advocates students’ idea of a predetermined ‘correct’ answer as evidenced by the phrase, ‘due to human error’, often used to explain unexpected results.

Much of what is taught in first year laboratory courses differs from, or contradicts, the current internationally recommended practice for professional scientists. The traditional approach to teaching measurement and handling data contains significant inconsistencies and does not take into account students’ existing views of measurement. This approach serves only to encourage students’ misconceptions about the nature of scientific measurement.

2.2 The new laboratory course

2.2.1 A probabilistic framework

The need for a consistent international language for calculating and communicating measurements and uncertainties prompted the Bureau International des Poids et Mesures (BIPM) to initiate a review in the 1970s, which culminated in the publication of a set of recommendations and guidelines issued by the International Organisation for Standardisation (ISO) in the 1990s. The two most widely known and authoritative publications are the International Vocabulary of Basic and General Terms in Metrology, *VIM* (ISO, 1993), and the Guide to the Expression of Uncertainty in Measurement, *GUM* (ISO, 1995). All international standards organisations including IUPAP (International Union of Pure and Applied Physics), IUPAC (International Union of Pure Applied Chemistry) and the U.S. National Institute of Standards and Technology (NIST) have adopted these recommendations. An abridged version of *GUM* is publicly available as

NIST Technical Note 1297 (Taylor and Kuyatt, 1994). The ISO guides are based on a probabilistic framework that follows the Laplace-Bayesian approach to statistical analysis.

In the probabilistic approach, in contrast to the frequentist framework, the data, or measurement readings, have no associated random error, but are constants. An inference about the quantity being measured, or measurand, is drawn from the finite data set and any pre-existing knowledge about the measurand. It is this inference about the measurand which has an associated uncertainty. Since all knowledge about the value of the measurand is contained in the inference, and additional data could modify this value, it is clear that the value of the measurand is a parameter which depends on the measurements performed. Consequently, information about the measurand is always incomplete and even the best inferred approximation must be accompanied by an estimate of how incomplete the information is, or a statement of the uncertainty.

This is in direct agreement with ISO specifications as detailed in paragraph 2.1 of TN1297: *"In general, the result of a measurement is only an approximation or estimate of the value of the specific quantity subject to measurement, that is, the **measurand**, and thus the result is complete only when accompanied by a quantitative statement of its uncertainty."* (Taylor and Kuyatt, 1994)

The uncertainty is defined as, *"a parameter associated with a measurement result, that characterizes the dispersion of the values that could reasonably be attributed to the measurand"* (GUM, ISO, 1995).

The measurement process involves the combining of prior knowledge about the measurand with new data to yield a best estimate of the measurand. The transition from making statements about the data to making statements about the measurand is enabled by Bayes' theorem and facilitated by probability theory (d'Agostini, 1999). Probability density functions (pdf's) are used to model the existing information and new data; they are then combined to form the final probability density function which encapsulates all knowledge about the measurand, and on which inferences about the measurand are based. This process is summarized in Figure 2.1.

The most common probability density functions used in metrology, as suggested by the ISO guide, are the Gaussian pdf, the rectangular (or uniform) pdf and the triangular pdf. For most practical measuring purposes the final pdf, if symmetrical (as are the most commonly used ones), is characterized by two quantities, the zeroth and second moments of the distribution (Campbell *et al.*, 2005). The zeroth moment coincides with the centre of the distribution and, in the context of measurement, the best estimate of the measurand. The second moment is related to the width of the pdf, and is called the variance. The standard uncertainty associated with the best estimate of the measurand is given by the square root of the variance. The interval described by $X \pm U$, where X is the best estimate and U is the standard uncertainty, is a measure of the incompleteness of knowledge about the value of the measurand. The aim of measurement in general, is to minimize this interval.

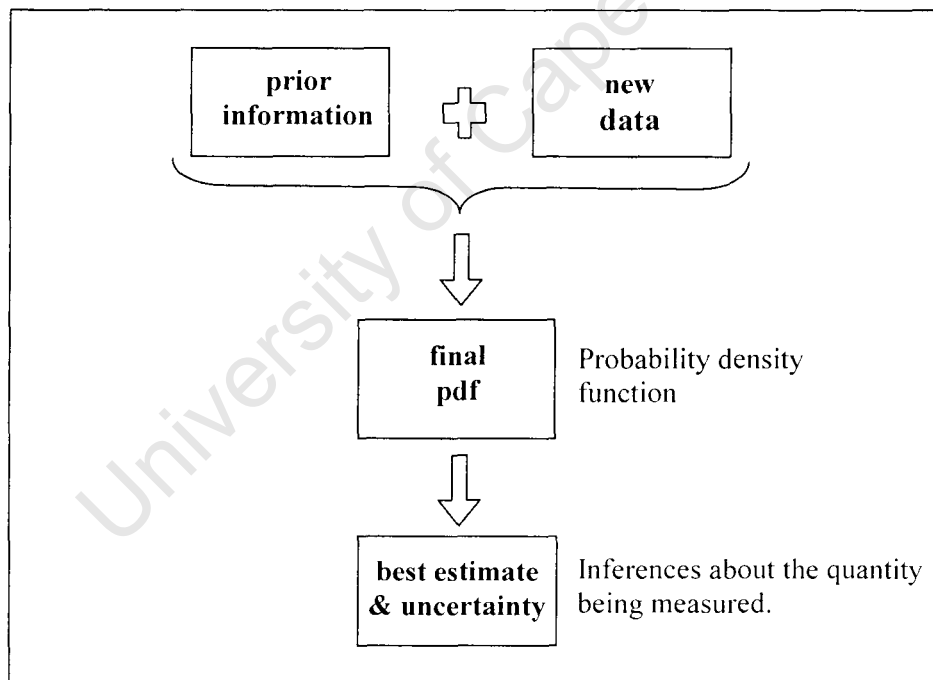


Figure 2.1: A model for determining the result of a measurement
(Allie *et al.*, 2004).

The area under the pdf spanned by $X - U$ and $X + U$ is a measure of the probability that the measurand lies within those limits, and is called the coverage probability or level of confidence. Each probability density function has an associated coverage probability: 68% for the Gaussian pdf, 65% for the triangular pdf and 58% for the rectangular pdf. A typical statement of a measurement result would take the following form: “The best estimate of the measurand is X with a standard uncertainty of U and the probability that the measurand lies on the interval $X \pm U$ is $Z\%$.”

According to the ISO guide (1995), uncertainty is classified into two types based on the method of evaluation. Type A evaluations are based on the use of statistical methods to evaluate the uncertainty associated with the dispersion of a data set, while Type B evaluations involve estimating uncertainty using available non-statistical information like instrument specifications, previous measurements and the observer’s judgement.

Practically, Type A evaluations of uncertainty are applicable to situations involving repeated observations with dispersion, and use Gaussian pdf’s and statistical formulae of similar structure to those employed in the frequentist approach but with different interpretations. Type B evaluations are applicable in all measurements and are usually modeled by triangular or rectangular pdf’s. This formalism allows a single measurement to be treated as easily and as consistently as a large sample of measurements.

It is important to note that the uncertainties resulting from Type A and Type B evaluations do not correspond to the random and systematic errors of the traditional scheme. This is clear in the treatment of systematic errors under the probabilistic formalism advocated by the ISO guide which states, “*Type B standard uncertainty is obtained from an assumed probability density function based on the degree of belief that an event will occur,*” implying that since systematic errors are never accurately known, they should acquire a probabilistic description. The overall uncertainty associated with a measurand is often a combination of uncertainties arising from both Type A and Type B evaluations. The ISO recommends an ‘uncertainty budget’ to calculate this value. The uncertainty budget is a list of all possible sources of uncertainty with an evaluation of each individual contribution based on the appropriate pdf. The combined uncertainty is then calculated using the familiar uncertainty propagation formulae. A notable feature of the uncertainty budget is

that any number of uncertainty components, whether resulting from Type A or Type B evaluations, can be included and combined as described.

This ISO endorsed probabilistic approach provides a logically consistent framework for teaching students the basic concepts of experimentation and measurement, and forms the basis of the new introductory physics laboratory course evaluated in this research project. A search for literature documenting the development and/or implementation of similarly - based physics laboratory courses proved unfruitful. This suggests that the course described here is the first, and only, of its kind currently being actualized as a part of an undergraduate physics curriculum.

2.2.2 Description of the new probabilistic laboratory course

The new laboratory course attempts to incorporate students' existing perceptions of measurement with the ISO's recommendations for evaluating and presenting scientific measurement, to achieve the intended learning outcome of the course: a coherent understanding of measurement and uncertainty. The course is run over sixteen weeks with one three-hour session per week. The sessions are alternately spent on written workbook exercises (see Buffler *et al.*, 2002) and 'hands-on' laboratory activities.

The laboratory activities are designed to support and elucidate the ideas introduced in the workbook exercises, with tasks presented as real-life scenarios in a style similar to that of the original course. Figure 2.2 presents the well-known pendulum practical recast in a conceivable real-world context.

The student workbook (Buffler *et al.*, 2002) is an interactive tool designed to introduce the concepts of measurement and uncertainty through exercises and activities, which challenge students' point perceptions of scientific measurement, and steer them towards adopting the set paradigm view of measurement and uncertainty. Students work through the activities in small groups in a tutorial-type environment with a number of roving tutors on hand to provide assistance when required. The broad content areas addressed in the workbook are listed in Table 2.1., and two example pages are presented in Figures 2.3 and 2.4.

Pendulum problem swings you into action

Imagine that you now work for a Scibucks Enterprises, a scientific company that consults for industry. Your boss calls you into her office and explains that she wants you to undertake an investigation for a client who is a clock maker. The clock maker says that he needs to know what the relationship is between the length of a pendulum and its period and must have evidence that this relationship works in practice. You remember from your undergraduate physics days that the period, T , of a pendulum is related to the length, l , of the string by

$$T = 2\pi \sqrt{\frac{l}{g}} \quad \text{where } g \text{ is the acceleration due to gravity.}$$

You therefore devise two experiments to test the theory:

Experiment A. Measure T for different lengths l and then plot a suitable graph to show that the above equation is valid.

Experiment B. Choose one length l and measure T many times, and then calculate g (using the equation above.) If your measured value $g \pm u(g)$ agrees with the theoretical value for Cape Town (9.79 m s^{-2}) then this would suggest that the equation for T is correct.

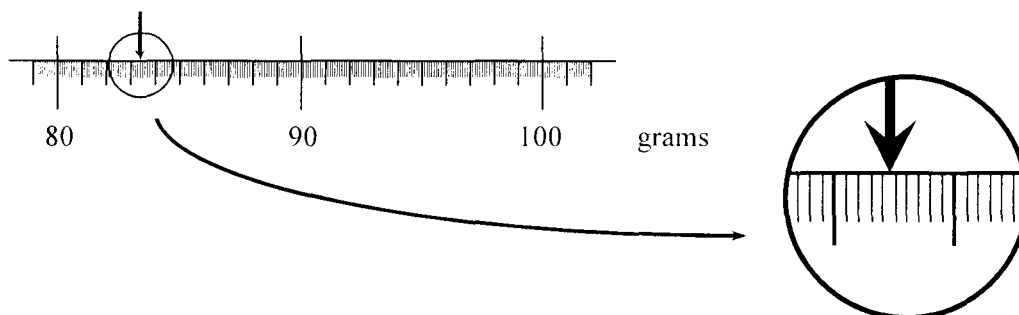
Your boss tells you that she must have your report completed before 10:00 on this Friday which should include a full description of your method, all the measurements you make, the calculations and graph, an uncertainty budget, and a suitable discussion and set of recommendations to the clock maker.

Figure 2.2: An example of an 'authentic' problem based practical exercise.

Table 2.1: Outline of the content of the interactive student workbook.

Unit	Description
1. Introduction to measurement	The relationship between science and experiment. The nature and purpose of measurement.
2. Basic concepts of measurement	Probability and inference. Reading digital and analogue scales. The nature of uncertainty. A probabilistic model of measurement.
3. The single reading	Probability density functions. Representing knowledge graphically using a pdf. Evaluating standard uncertainties for a single reading. The result of a measurement.
4. Repeated readings that are dispersed	Dispersion in data sets. Evaluating standard uncertainties for multiple readings. Type A and type B evaluation of uncertainties.
5. Working with uncertainties	Propagation of uncertainties. Combined standard uncertainty. The uncertainty budget. Comparing different results. Repeatability and reproducibility.
6. Modelling trends in data	Principle of least squares. Least squares fitting of straight lines.

If you want to subdivide each graduation even further, you now have a division marker every 0.1 g. You might need a magnifying glass to read the scale !



The reading on the scale is : _____



It becomes impractical to continue to add more and more subdivisions. Eventually the scale becomes too small to read. No matter what analogue scale you are reading, you will always need to

Will you ever be able to find an instrument that gives you a reading of the mass of the block to an **infinite** number of decimal places? No, of course not. It will never be possible to manufacture such an instrument ! It is then clear that the "true" value of the mass can never be known. This is the case for **all** measurements, no matter what you are wanting to measure.

Figure 2.3: An example page from the student workbook (Buffler et al., 2002).

Actually, it is usually not possible to identify a single reason for what causes the observed scatter in the data. Even if you do the experiment as carefully as possible, then there will still be a dispersion in the readings of d . The important question is how to deal with this dispersion (in this case d).

So far we have 4 rolls from the same height and have determined :

$$d_1 = 650.4 \text{ mm} , d_2 = 660.6 \text{ mm} , d_3 = 659.1 \text{ mm} \quad \text{and} \quad d_4 = 669.6 \text{ mm}$$

The best approximation for d after one roll is clearly 650.4 mm. After 2 or more rolls, the **average**, or **arithmetic mean**, of all the readings is usually the best value to use.

Why is this the case ?

After 2 rolls, the average is _____ mm

After 3 rolls, the average is _____ mm

After 4 rolls, the average is _____ mm

You can see that the average changes as we take more and more readings.

Now I am really confused ! **How many** readings should I take when doing an experiment and my data are showing a scatter ?



It is not possible to give a firm answer to this question. Let us say that you decide to roll the ball a total of 50 times from the same height, $h = 78.0 \text{ mm}$. Then you might see the following pattern of spots on the paper:

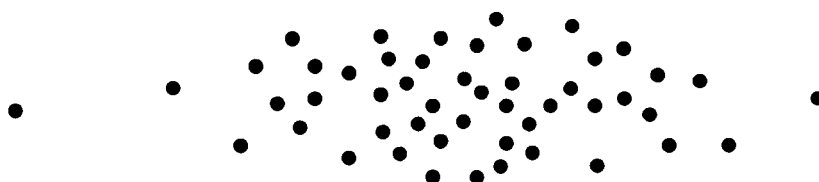


Figure 2.4: Another example page from the student workbook (Buffler et al., 2002).

The introduction to measurement begins with leading students to the discovery of the concept of a measurand, and to the realization that a measurement always involves a comparison with a reference standard. Various measurement exercises are then used to illustrate the fact that the reference standard can never be infinitely small, thus introducing the idea that the knowledge obtained about a measurand through the measurement process is always limited.

After an exploration of the different purposes of measurement in everyday and scientific contexts, the exercises move on to deal explicitly with a core component of the course - the difference between the reading observed on the measuring instrument and the conclusions that can be drawn about the value of the measurand. The final introductory exercise requires students to reflect on a previous experiment and to list all the factors likely to have influenced their measurement results, together with an indication of the relative magnitudes of the effects. For example, the air temperature on the day influenced the results in a “small” way and the student’s skill with the apparatus had a relatively “large” effect. This highlights the need for a universally meaningful way to analyse and communicate measurement results, and provides the motivation for subsequent sections and the course as a whole.

The next section introduces the concept of probability and then investigates the uncertainties associated with reading both analogue and digital scales. The focus is on what can be inferred about a measurand from a single digital or analogue reading. Given a reading on a digital scale sensitive to one decimal place, students are asked to predict what the second decimal digit will be if the sensitivity of the scale is increased by a factor of ten. The majority of students are quick to realize that there is an equal probability of the unknown digit being any whole number from 0 to 9. Further examples lead them to conclude that the digital scale can theoretically be made ‘infinitely sensitive’ to give a reading with infinitely many decimal places, while also appreciating the practical impossibility of creating such an instrument.

Students are then required to consider readings on analogue scales with increasingly fine graduations. The obvious conclusion in this case is that reading the scale is dependent on the observer’s judgement. Students are forced to concede that even in the absence of all

outside sources of uncertainty, the knowledge about the value of the measurand will always be limited to an interval, the width of which can never be reduced to zero. This challenges students' widely held belief in the possibility of knowing the "true value" of a measurand.

Once students have assimilated the fact that a measurement result in science is meaningless without a quantitative statement of the uncertainty, they are introduced to the formal tools for dealing with uncertainties - probability density functions. Students are prompted to recall the definitions of density and density functions before being presented with the new idea of a probability density function. The most commonly used pdfs in metrology, (the Gaussian, the rectangular pdf and the triangular pdf), are listed along with flexible guidelines for their use, but the initial focus is on the rectangular and triangular pdf's. These probability density functions are used for Type B evaluations of uncertainty, which are characterized by the use of available knowledge of the measurement process and/or apparatus to determine some measure of the uncertainty. Examples of Type B uncertainties are the uncertainty associated with reading a scale, and the uncertainty associated with the internal calibration of a measuring instrument. Various exercises involving single readings on both digital and analogue scales are employed to illustrate that the pdf summarizes all available information about the measurand, with the best approximation, or most likely value, of the measurand corresponding to the center of the pdf, and the standard uncertainty related to the average width of the pdf. Additional exercises deal with reporting the result of a measurement as probabilistic statement, and reinforce the fundamental set paradigm tenet that the measurement process involves modeling all new data with existing information to yield a measurement result.

Handling scatter in a set of repeated observations of the same measurand (Type A evaluation of uncertainty) is deliberately delayed till students are adequately able to deal with a single measurement. This is necessary to dispel the pervasive belief among university entrants that dispersion in data is the dominant source of uncertainty and that the average value accounts for all "experimental errors". The intention is to introduce dispersion as one of many sources of uncertainty.

An experimental context is presented, in which a ball is released from a particular height on a ramp attached to a table, rolls down the ramp, and onto a sheet of paper on the floor, leaving a spot to mark where it lands. The ball is released from the same height a few times, and each roll produces a spot on the floor in a slightly different position to the one before. The student is faced with many questions: How many rolls is enough? How can the scatter best be modeled? What is the most likely value of the measurand?

To help the student answer the questions, a table of 50 readings with a running average is given. The student's attention is drawn to the fact that the average stabilizes as the number of readings increases. With the aid of relative frequencies, histograms and curve fitting the student is led to the conclusion that the Gaussian pdf is the most appropriate for modeling the given data. The best approximation of the measurand corresponds to the centre of the pdf and the mean or average value of the data, and the standard uncertainty is related to the average width of the Gaussian and the standard deviation of the mean. The statistical formulae for the mean and standard deviation are then introduced, and used to perform Type A evaluations of uncertainty in given examples.

Having successfully exposed students to the methods of dealing with a range of sources of uncertainty, the next few chapters guide them through propagating uncertainties, determining the combined standard uncertainty and comparing results. The pervading theme of considering and evaluating every possible source of uncertainty culminates, at the end of this section, in the concept of an "uncertainty budget" which enables the student to determine a reasonable total uncertainty for a particular measurement.

A final chapter tackles the principle of least squares and least squares straight line fits through a combination of theoretical and practical exercises. The workbook is peppered throughout with practical examples requiring students to engage with various apparatus and obtain and analyse data.

Appendices in the workbook cover many pertinent subjects including guidelines for planning experiments, writing laboratory reports, drawing up tables and plotting graphs, detailed notes on probability density functions and expanded uncertainties, instructions for

the use of certain measuring apparatus and the statistical functions on scientific calculators, and exercises on converting units.

This laboratory course was piloted in 2002, with 160 first year students in the Physics Department at UCT. A modified and improved version has been run in all subsequent years, to date.

University of Cape Town

Chapter 3: Method

3.1 Description of the sample

This study is an evaluation of the probabilistic laboratory course taken by first year science students registered for General Entry to Programmes in Science (GEPS) at the University of Cape Town (UCT), who have physics as one of their courses. GEPS is a structured year-long bridging BSc programme which extends the mainstream BSc degree to four years instead of nominal three. It is targeted mainly at educationally disadvantaged students, who fail to meet the minimum entrance requirements for the mainstream, or 'direct entry', programme. Selection for admission to GEPS is based on factors including secondary school examination results, equity quotas, and various other indicators of potential for success.

The majority of students in GEPS come from schools which, prior to May 1994, were run by the now defunct Department of Education and Training, (the administrative body governing black schools during the apartheid era in South Africa), and are in areas which are severely economically challenged. As a result, the background of a typical GEPS student is disadvantaged not only educationally, but also socially and economically. An additional challenge is communication in English, which most students do not speak as a first language.

The South African education system is in the process of reconstruction. While the significant variation in quality of education at different schools is being addressed, the effects are understandably still evident. Consequently, many students entering science programmes in higher education have their first experience of laboratory work in the first year laboratory.

The target sample included approximately 150 GEPS students, 100 of whom volunteered to participate in the pre-intervention survey and 117 in the post-intervention survey. Of these, 76 completed both the pre- and post-intervention questionnaires.

3.2 Research methods

The data used in this study were obtained from survey-based investigations with the primary data coming from written questionnaires, and one-on-one interviews providing additional information. A combination of simple quantitative, and more complex qualitative, research methods were employed because the aim of this study is to explore students' ideas and gauge their understanding. Among the challenges associated with conducting surveys of this nature is making the questions accessible to all respondents in terms of context and phrasing.

The individual shortcomings of the written questionnaire and the interview also need to be considered, and subsequent discussion will illustrate how these were addressed and also how certain benefits of the one method compensated for disadvantages of the other.

The questionnaire is a research tool which requires the respondent to answer a posed question either by choosing from given options (closed questions), or by offering an independent answer (open questions), or picking from a list of actions and providing motivation for the choice (semi-structured questions). The questionnaire has the advantages to the researcher, of being easy to construct, relatively cost efficient in terms of both time and money, and viable for very large sample groups. The respondent is free from the strain of having to offer an immediate response and is given time to think. Sources of bias in the data acquiring process are minimized by the standardizing of questions, and the absence of an interviewer

Gillham (2000) points out that this does not, however, take into account the levels of literacy and writing skills of the respondents, nor does it allow the probing of responses to elicit further explanations, the clarification of misunderstandings, or the gauging of the sincerity of responses. Other concerns highlighted include motivating the students to participate in the survey, and having achieved this, engaging and maintaining their interest through the questionnaire, and ensuring that the order in which questions are answered is the same as that in which they're asked.

These latter problems are largely eliminated by careful attention to the design of the questionnaire and to the conditions under which it is administered. Where responses require clarification or explanation, one-on-one verification interviews are employed to improve the validity of the analysis.

Cohen *et al.* (2000) discuss the pros and cons of the one-on-one interview. While time-consuming for both data collection and data analysis, the interview provides an opportunity for the researcher to delve into the students' responses and investigate the rationale and motivation behind them. Students are given the opportunity to express themselves more effectively, and data is thus more detailed and complete, but it is also more open to the prejudice and interpretation of the interviewer. Another downfall of conducting a survey by interviews only, is the restriction that the time expense places on the sample size.

With these factors in mind, it was clear that the most effective way to acquire the required data was to administer a semi-structured open questionnaire to the sample group, and follow up with one-on-one verification interviews as required.

3.3 Design of the questionnaire

3.3.1 Background of questionnaire design

Designing the probes was not an explicit part of this research project, but the process is described here for completeness.

The template for the design of the questionnaires used in this study, and in previous ones also forming part of the UCT-York collaborative research project (Allie *et al.*, 1998; Buffler *et al.*, 2001; Lubben *et al.*, 2001), was provided by the instruments developed for the PACKS project (Lubben and Millar, 1996). The original PACKS questions, or 'probes', could not be used directly in the studies conducted at UCT, since they presented multifarious contexts relevant to, and specifically aimed at, the 11 - 15 year old UK school children on which the PACKS project focused.

Therefore, using the PACKS instruments as a guide, new probes were developed for the university studies. To address the previously mentioned disadvantages associated with questionnaire type surveys, as well as the additional concerns of respondents' often apparent difficulty visualizing hypothetical situations, and of finding a single context relevant to the different sample groups surveyed in the project, two simple first year laboratory experiments were chosen as the bases for the probes.

3.3.2 Experimental scenarios

The experimental scenarios were chosen primarily for their lack of complicated detail, which made them easy to describe and to visualize. Another criterion for their selection was the fact that the students were unlikely to have encountered the experiments prior to seeing the questionnaires; this would ensure that the responses to the probes would not be based on rote from prior experiences.

The pre-intervention scenario describes an experiment in the laboratory using a sloping wooden ramp clamped to the edge of a table, a small metal ball, a metre rule and marking paper. The small metal ball is released from various positions along the sloping ramp; it rolls down the ramp and onto the marking paper affixed to the floor. The situation is clearly explained, with an accompanying diagram, on the cover sheet of the questionnaire, which is reproduced in Figure 3.1.

To make sure that the context was completely understood, the experiment was demonstrated using a 'life-sized' model of the wooden ramp and a tennis ball. The ball was released from two different positions on the slope to illustrate, without additional comment which could potentially compromise the responses to the probes, how the distance d changes with the height h .

An experiment is being performed by students in the Physics Laboratory. A wooden slope is clamped near the edge of a table. A ball is released from a height h above the table as shown in the diagram. The ball leaves the slope horizontally and lands on the floor a distance d from the edge of the table. Special paper is placed on the floor on which the ball makes a small mark when it lands.

The students have been asked to investigate how the distance d on the floor changes when the height h is varied. A metre stick is used to measure d and h .

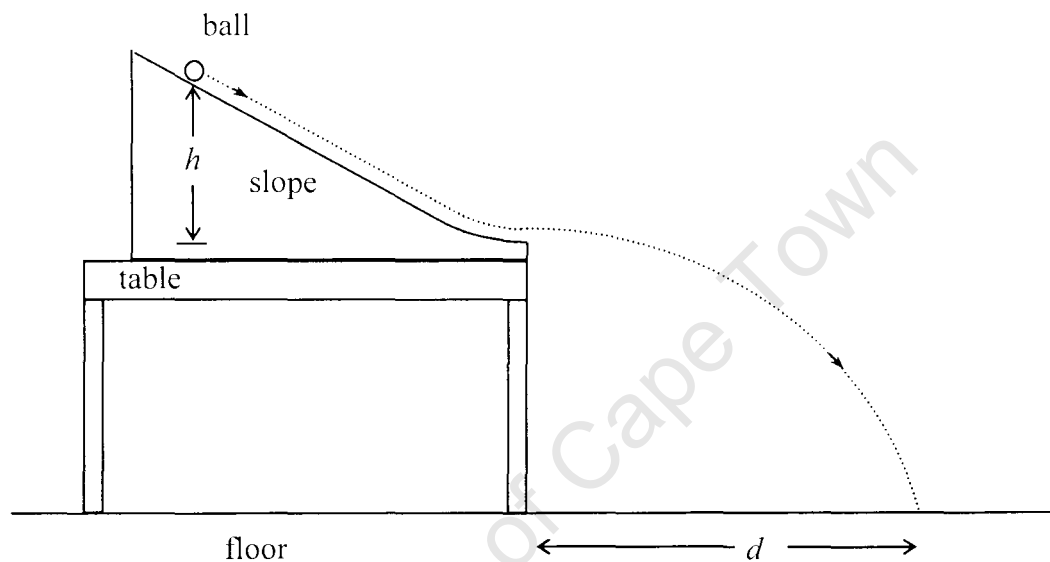


Figure 3.1: Description of experimental context as it appears on pre-intervention questionnaire covers.

Students encounter this experiment again during the year as part of the physics laboratory course. A second experimental context is thus chosen for the post-intervention questionnaire.

The second scenario describes an experiment requiring a compressed spring and a small block on a table with non-negligible friction and a metre rule. The block is pushed against the spring and then released. It travels a distance d which is determined using the metre rule. As with the pre-test scenario, a brief but clear description of this experiment with an illustration of the situation is provided on the questionnaire cover page, as shown in Figure 3.2, and a scaled-up version of the apparatus is used to demonstrate the experiment.

An experiment is taking place in the physics laboratory to investigate the motion of a block on a table with friction. The block is pushed against the spring so that its left edge is at position P . The block is released and travels a distance d to position Q as shown. The students, working in groups, have to determine d using a metre rule that is provided.

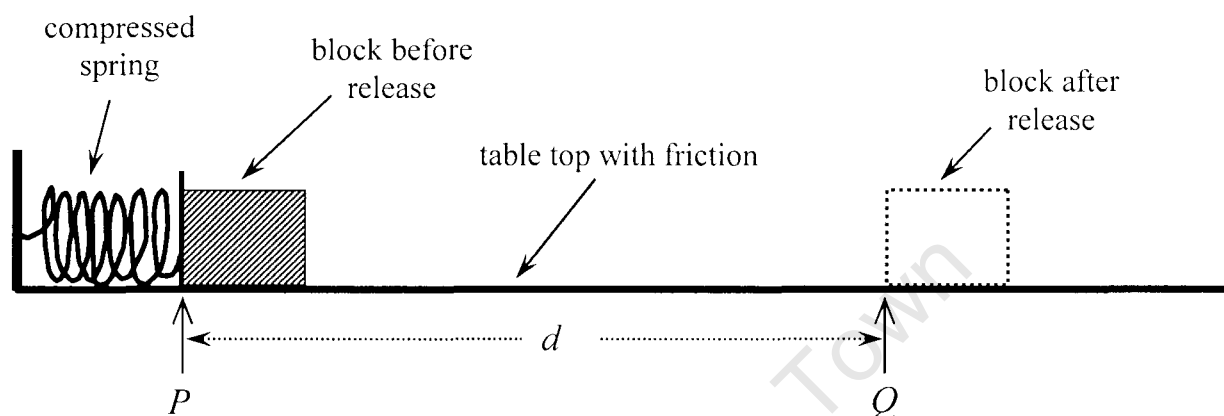


Figure 3.2: Description of experimental context as it appears on post-intervention questionnaire covers.

3.3.3 Questionnaire probes

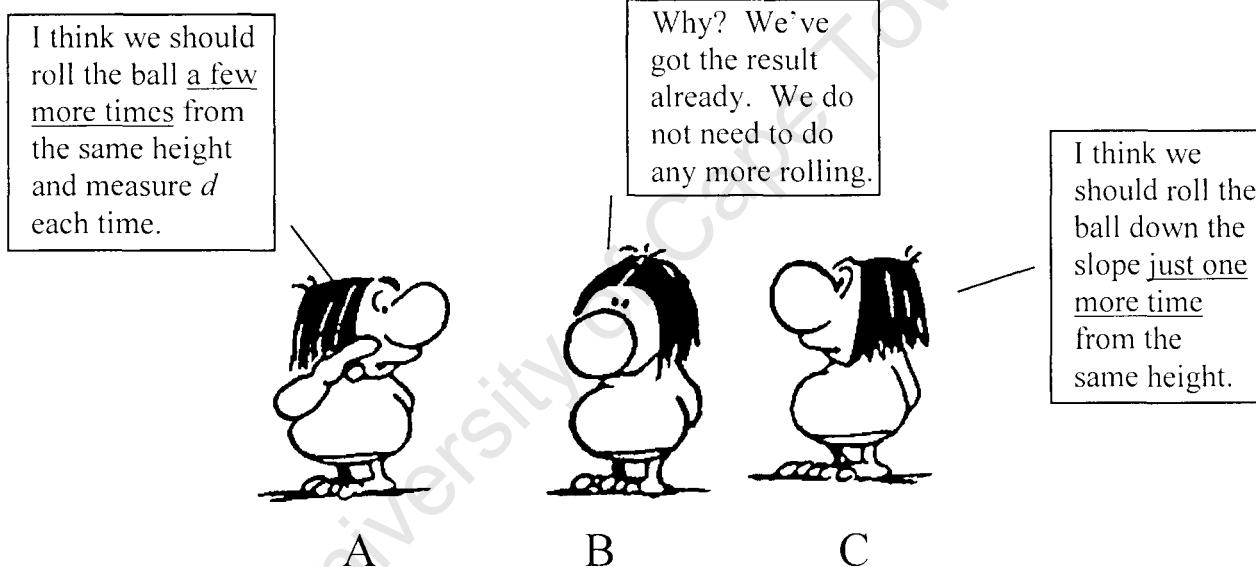
The individual probes were all similarly structured and each of the two sets of probes was based on a given experimental scenario. A brief description of a practical laboratory situation was given, and a course of action had to be decided on. Carefully chosen cartoon characters presented the various options available. The student was required to make a choice and then provide a justification for the choice (Campbell *et al.*, 2005). A typical probe is illustrated in Figure 3.3.

The nature of this study and its focus on investigating students' ideas and understanding of measurement and uncertainty, ruled out the multiple choice questionnaire format, since the respondent would then be limited to choosing from the researchers' ideas with no opportunity to explain their choices, and no recourse should they have completely different views. The semi-structured format allows the researcher to guide the respondent

by offering choices relevant to the particular area probed, while still leaving room for explanations. This limits the range of responses and simplifies analysis. Care was taken to ensure that the possible actions presented could each be chosen for a variety of reasons. Most probes also offered the choice for the respondent to suggest an alternative action. These were important considerations for limiting responses made merely by recognition rather than by understanding.

The students work in groups on the experiment. Their first task is to determine d when $h = 400$ mm. One group releases the ball down the slope at a height $h = 400$ mm and, using a metre stick, they measure d to be 436 mm.

The following discussion then takes place between the students.



With whom do you most closely agree? (Circle ONE):

A	B	C
---	---	---

Explain your choice.

Figure 3.3: The RD (Repeating Distance) probe.

Since the majority of students in this study did not have English as a first language, it was important to keep the vocabulary and language structure as simple as possible. The probes were thus phrased in a concise, terse manner with the appropriate terminology included in the text. The cartoon characters, (from Geoff Watson's "King Tut"), were chosen for their racial, ethnic and gender anonymity. This was done to eliminate the possible influences of true-to-life images, with names and cultural identities, on the students' responses (Allie *et al.*, 1998).

The decision to adopt the cartoon characters labeled with letters of the alphabet was made after the design was tested by presenting it, together with an alternative using life-like characters, or 'talking heads', with names and easily recognizable race and gender, to a sample of the target group, and soliciting their opinions through interviews and written feedback. The vast majority of interviewees preferred the anonymous cartoon characters, and confirmed the accessibility of the vocabulary and text composition (Allie *et al.*, 1998).

3.4 The questionnaires

Two questionnaires were administered to the target cohort of students during the year: the first in February before the start of the GEPS course, and the second in September, at the end of the course. The questionnaires were both composed of a series of page-long questions that required written answers and that were based on the experimental contexts described later.

The first (pre-intervention) questionnaire, or 'pre-test', comprised thirteen probes (see Appendix I), and the second (post-intervention) questionnaire, or 'post-test', consisted of nine probes (see Appendix III), five of which were the same as ones in the pre-test, and four which were similar, but not identical, to pre-test probes. Seven of the seventeen different probes (RD, RDA, SMDS, DMSS, UR, NU1 and NU2) used in this study were used in previous studies (Allie *et al.*, 1998; Buffler *et al.*, 2001) and have thus been validated. Ten probes (UA1, UA2, QD1, QD2, QD3, PR1, PR2, ED1, PX1 and AE1) were newly developed, or modified from existing probes, exclusively for this investigation.

Each probe targets a particular aspect of measurement and the students' decisions in each case reveal their understanding of the specific area. The RD and RDA probes are concerned with data collection with a focus on repetition of measurements, and the SMDS and DMSS probes deal with data comparison. These probes together with the data processing probe which solicits students' ideas about the best representation of a data set, the UR probe, have been included in many previous studies and extensively analysed. The newly developed QD1, QD2, QD3, PR1, PR2, ED1, UA1 and UA2 probes focus on data processing. The UA probes deal with averages of sets of measurements, and the QD1, QD2, QD3, PR1, PR2 and ED1 probes are concerned with single measurements. The NU1, NU2 and PX1 probes explore students' views about uncertainty and the AE1 probe solicits students' concepts of the nature of measurement.

The probe pairs, QD1 and PR2, QD2 and PR1, and QD3 and ED1, each explore the same ideas about single measurements, in different experimental contexts. For example, the QD3 and ED1 probes both deal with a single measurement on a digital scale, with the QD3 probe framing the question in the context of the rolling ball experiment, and the ED1 probe set in the context of a block and spring experiment. Similarly, the NU1 and PX1 probe share the underlying concept of the reducibility of uncertainty, with the NU1 probe proposing that efficiency in experimental practice is the route to measurements with zero uncertainty, and the PX1 probe offering the perfect measuring apparatus as the means to the same.

The inclusion of more than one probe into the same aspect of measurement - for example, both the RD and RDA probes deal with repeating measurements, and the UA1 and UA2 probes are both focus on averages of data sets - increase the reliability of responses obtained, i.e. consistent or inconsistent responses indicate whether or not the questions are understood, and the responses carefully considered.

The complete list of probes and their descriptions is given in Table 3.1.

Table 3.1: *List of probes used in this study.*

Aspect of measurement	Pre-test probes	Post-test probes
Data collection	RD - Repeating Distance RDA - Repeating Distance Again	- -
Data processing	UR - Using Repeats UA1 UA2 QD1 QD2 QD3	UR - Using Repeats UA1 UA2 PR2 PR1 ED1
Data comparison	SMDS - Same Mean Different Spread DMSS - Different Mean Same Spread	SMDS - Same Mean Different Spread DMSS - Different Mean Same Spread
Uncertainty	NU1 - No Uncertainty 1 NU2 - No Uncertainty 2	PX1 -
Nature of measurement	AE1	-

Since students are required to answer the questions in a strict sequence, a final probe included in both the pre- and post- intervention questionnaires affords them opportunity to amend earlier responses and to make comments.

3.5 Administration of the probes

Students answered the questionnaires under examination conditions on a voluntary basis. It was made clear to them that their participation was not compulsory and would not have any impact on their course evaluation. It was also stressed that the information provided would be used for research purposes only and that although they were required to write their names and/or student numbers on the questionnaire, this was necessary only for linking the pre- and post- intervention responses.

Each questionnaire was given an alphabetic code which identified it among the various questionnaires used in the various studies in the larger research project. The pre- and post-questionnaires used in this study were assigned the letters K and L respectively.

Each probe was printed on an A4-sized sheet of paper and labeled at the top of the page with a question number, an abbreviation of the code description and the questionnaire letter code; for example the Repeating Distance probe, question 2 of the pre-intervention questionnaire, was labeled Q2. (RD/K). The numbered probes were put into a brown A4-sized envelope, and the cover sheet was pasted to the front of the envelope.

A unique number was stamped in a box below the one provided for the student's name on the cover sheet, and reproduced on all the probes in the specific set. This laborious process had a dual purpose: it helped the researcher to remain impartial during analysis, while still being able to identify the probe responses associated with each student, and it facilitated the comparison of the individual probe responses.

Participating students were each given an envelope containing the ordered probe sheets, with a list of instructions and a description of the experimental context pasted on the front. The administrator then read the instructions to the students, taking care to make sure that they were clearly understood, and answering any questions posed. It was emphasized that there were no right or wrong answers to the probes, and that clear, detailed explanations were of paramount importance. About a third of each probe sheet was reserved for free responses, and students were encouraged to use the reverse sides of the sheets should they require additional space.

They were also urged to answer the questions in the set order, and not to skip any. This strict sequence was necessary since latter probes were likely to influence the responses in earlier ones. Maintaining the given sequence ensured the least contaminated responses. In an attempt to enforce this, students were told to put completed probes back in the envelopes and not take them out again.

Since the analysis included probe-by-probe coding, it was important for each response to be self-contained. Students were thus asked to write full answers to all the questions even if they were similar to, or exactly the same as, previous answers. They were then reminded that the last question would give them the opportunity to modify or change any responses. There was no strict time limit given, but it was suggested that students spend approximately five minutes on each question to gain a suitable grasp of the situations presented before making decisions.

The rules governing the administration of the probes, including the strict answering sequence and examination conditions, and the experiment demonstrations served to increase the reliability of the data collected.

3.6 Analysis method

3.6.1 Construction of coding schemes

The completed responses to each probe were collected, and to facilitate the coding process, sorted according to the set numbers stamped on the envelope, and on each sheet. Each probe required a coding scheme; the probes developed for previous studies already had schemes, some probes required the modification of existing coding schemes, while others required the development of new schemes.

The Grounded Theory method (Strauss and Corbin, 1990) was used to develop the coding categories. The individual responses to each probe were read carefully to identify the main ideas and classify them. This was done independently by members of the research team, who then came together to compare, clarify, refine and label each category with an appropriate descriptor. To ensure that all ideas elicited from the responses were clearly represented in the classification scheme, categories were subdivided to take into account subtle distinctions between different responses. Each response was given an alphanumeric code based on the student's choice of action and supporting explanation. For example in the RD probe (Figure 3.3) the respondents' reasoning for the choice of action A, (roll the ball a few more times from the same height), included the idea that repetition is required to

practice the experimental procedure; this idea was categorized as A1 with the descriptor '*practice makes perfect*'. Further investigation revealed more explicit reasons for practicing the experiment. The A1's were then subdivided to include all these reasons: A10 - *practice will produce a more accurate or better measurement*, A11 - *practice will reduce the systematic error in the measurement*.

To facilitate the analysis of responses in terms of point and set paradigms, each descriptor was compared to the definitions of the point paradigm and the set paradigm, (Table 1.1). Descriptors matching the definition of the point paradigm were assigned the paradigm code P, and those matching the definition of the set paradigm were assigned the paradigm code S. Where there was significant ambiguity, the paradigm code U was assigned.

These draft coding schemes were used by two more researchers, (who had not been involved in the development of the schemes), to classify a particular set of responses. Comparison of the classifications of these two researchers provided a measure of the validity of the scheme and a suggestion of necessary amendments. This procedure was iterated until the scheme stabilized (any differences were resolved by considering all the responses of an individual student across sets of related probes). The full coding schemes for all the probes are presented in Appendices II and IV.

3.6.2 Reliability of code allocation

In order to ensure the reliability of the code allocation, three researchers independently coded at least twenty responses to each of a few selected probes using the consolidated coding schemes. The allocated codes were then compared and the level of inter-coder agreement was established.

3.6.3 Individual probe analysis

An alphanumeric code is allocated to each response as described earlier. The codes for each probe are then grouped according to the main ideas evidenced by the students' responses. The resulting categories are classified as point (P), set (S) or indeterminate (U) paradigm-based, and the frequencies of responses in each category are calculated.

3.6.4 Cross probe analysis

For the cross probe analysis, the probes are grouped according to the aspect of measurement they address - data processing, data comparison, and uncertainty in measurement - to facilitate the evaluation of students' use of the point and set paradigms in decision-making at different stages of performing an experiment.

Paradigm codes are assigned to each student's set of responses to a particular probe group. The paradigm code S is assigned where the set paradigm is consistently employed; the paradigm code P indicates consistent use of the point paradigm; M represents mixed paradigm use and is assigned where the responses are neither consistently point- nor consistently set-based; and the paradigm code U is assigned where the paradigms used are ambiguous or cannot be determined. The number of students falling into each paradigm category, before and after the intervention, is tabulated for each probe group and across all probes.

3.7 Interviews

After students completed the post-intervention questionnaire and once the responses had been classified according to the developed coding schemes, a sample of thirty volunteers from the group of respondents was interviewed by a researcher. The interviews were one-on-one and each was approximately thirty minutes long. The purpose of the interviews was manifold: the researcher's interpretation of the student's responses was compared to the student's intended meaning in order to validate the coding schemes, the student's understanding of the questions was checked, and the validity of the probes, in terms of their accessibility to the students, was assessed.

The general impression was that the questions were clearly presented and understood, and that the probes offered a wide enough range of possible actions to encompass all the ideas expressed by the respondents. It was found that the researcher's interpretation of responses was mostly consistent with student's ideas, and consequently, that the coding schemes were suitably valid.

Chapter 4: Findings - Individual Probe Analysis

Abridged versions of each probe along with the tabulated frequencies of responses in each category and illustrative quotes from students' responses are presented below. The RD, RDA, UR, SMDS and DMSS probes have been extensively analysed in previous studies; the responses in this study express views similar to those extracted in the earlier studies, and are thus only briefly summarized here. (The main ideas are tabulated, but no illustrative quotes are provided.) Also excluded from this analysis is the NU2 probe which failed to elicit any ideas different from those expressed in the responses to the NU1 probe.

The data are presented as pre-post comparisons, with the exceptions of that from the RD, RDA and AE1 probes which were administered only in the pre-intervention questionnaire; Also, the data from the NU1 and PX1 probes are not compared, since, although they both deal with uncertainty in measurement, they focus on different aspects of the experimental process, viz. experimental skill and experimental apparatus, respectively.

4.1 Data collection probes: Students' ideas about repeating measurements

4.1.1 The RD and RDA probes

The RD and RDA probes are concerned with whether or not repetition of measurements is necessary.

The RD (Repeating Distance) probe:

The following discussion takes place between the students.

- A:** I think that we should roll the ball a few more times from the same height and measure d each time.
- B:** Why? We measured d already. We do not need to do any more rolling.
- C:** I think we should roll the ball down the slope just one more time from the same height and measure d again.

The RDA (Repeating Distance Again) probe:

After two rolls from the same height of $h = 90$ mm, the students have the following readings:

First release: $h = 40$ mm $d = 436$ mm

Second release: $h = 40$ mm $d = 426$ mm

The following discussion then takes place between the students.

- A:** We know enough. We don't need to roll the ball again.
- B:** We need to roll the ball just one more time.
- C:** Three rolls will not be enough. We should roll the ball several more times and measure d each time.

An overview of the respondents' views on repeating measurements when collecting data is presented in Table 4.1.

Table 4.1: *Students' paradigm use in their responses to the pre-intervention RD and RDA probes (n = 100).*

Category	Description	Codes		Paradigm	No. of students (%)	
		RD	RDA		RD	RDA
R1	No repeats are necessary.	B30; B40; B50	A40	P	2 (2)	1 (1)
R2	Repeats are a waste of time and resources.	C50; C51	A50; B40; B50; B51	P	0 (0)	2 (2)
R3	Repeats are necessary (no reason).	A40; A60; A62; C40	B60; C40; C60	P	8 (8)	11 (11)
R4	Repeats provide practice for improving the measurement.	A10; A11; C10; C11	B10; B11; C10; C11	P	8 (8)	8 (8)
R5	Repeats are necessary for improved accuracy.	A12; A64; A74	B64; B74; C12; C64; C74	P	8 (8)	17 (17)
R6	Repeats are necessary for finding a recurring value/pattern.	A30; C30	B30; B31; C30; C31	P	50 (50)	26 (26)
R7	Repeats are necessary for finding the mean/average.	A20; A21; A22; A23; A24; C20; C21	B20; B21; B22; B23; B24; C20; C21; C22; C23; C24	S	22 (22)	33 (33)
R0	Uncodeable responses.	U00; A01; B01; C01		U	2 (2)	2 (2)

The majority of respondents (96% in RD, 95% in RDA) believe that repeating measurements *is* necessary. The predominant reason given in the RD responses is that repeats will reveal a recurring value or a pattern in the readings, while the RDA probe sees responses split fairly evenly between repeating to find a recurring value and repeating to calculate the average. While the latter category indicates some set-based thinking, the reasoning overall is clearly based on the point paradigm, as reflected in 76% of the RD responses and 65% of the RDA responses.

4.2 Data processing probes: Students' ideas about single measurements

The QD1, QD2, QD3, PR1, PR2 and ED1 probes attempt to elicit students' views on the relationship between a single measurement and the quantity being measured (measurand). The QD probes were included in the pre-intervention questionnaire and the PR and ED probes in the post-intervention questionnaire. The pre and post probes explore the same ideas in different experimental contexts.

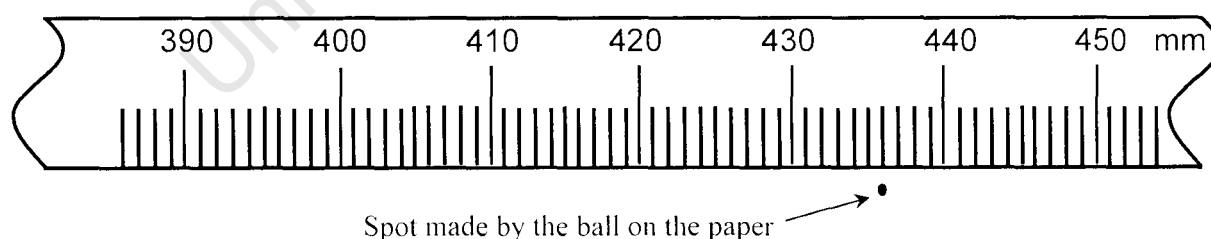
4.2.1 The QD1 and PR2 probes

The QD1 and PR2 probes focus on a single reading on an analogue scale, with the reference point directly below a graduation.

The QD1 probe:

The text of the QD1 probe reads as follows:

The students work in groups on the experiment. Their first task is to determine d when $h = 90$ mm. One group lets the ball roll down from a height $h = 90$ mm and use a metre rule to measure the distance d . What they see is shown below:

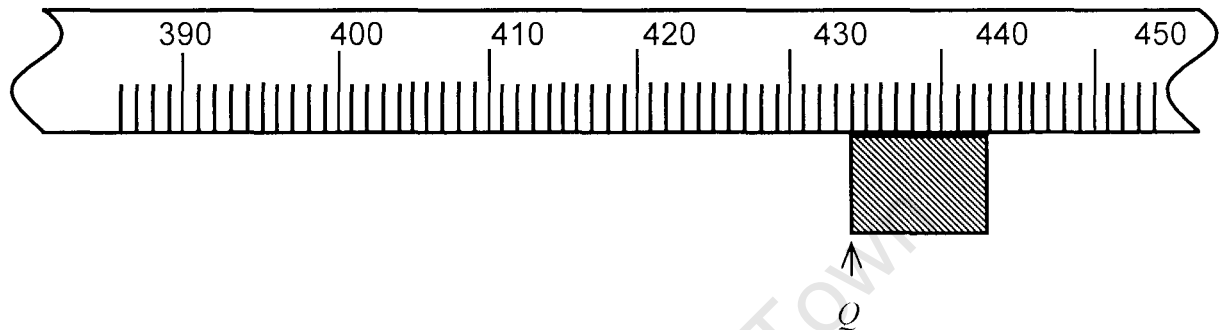


- A: I think the distance d is exactly 436 mm.
- B: I think the distance d is approximately 436 mm.
- C: I think the distance d is between 435 mm and 437 mm.
- D: I think the distance d is between 435.5 mm and 436.5 mm.
- E: I don't agree with any of you.

The PR2 probe:

The PR2 probe frames the question in a different context:

A second group of students marks point P and carefully lines it up with the zero mark on the metre rule. They then release the block. After the block comes to rest they see that point Q lines up on the metre stick as shown.



The students then have the following discussion.

- A:** I think the distance d the block has travelled is exactly 434.0 mm.
- B:** I think the distance d the block has travelled is approximately 434.0 mm.
- C:** I think the distance d the block has travelled is between 433.0 mm and 435.0 mm.
- D:** I think the distance d the block has travelled is exactly 434 mm.
- E:** I don't agree with any of you.

The common intent of the QD1 and PR2 probes, and the corresponding ideas identified among students' responses to each, facilitated a pre and post comparison of response frequencies. Table 4.2 contains the details of the comparison.

Table 4.2: *Students' paradigm use in their pre- and post-intervention responses to probes dealing with a single reading on an analogue scale (reference point appears directly below calibration line) (n = 76).*

Category	Description	Codes	Paradigm	No. of students (%)	
				Pre	Post
SRO1	Measurement is exact.	A20	P	4 (5)	0 (0)
SRO2	Repeats will give recurring value.	A41; B31	P	1 (1)	0 (0)
SRO3	Better measuring device/more sensitive scale will give accurate/exact results.	A12; A13; B12; B13; B15; C12; D12; D13	P	22 (29)	14 (18)
SRO4	The size, shape and movement of the ball/block affect the reading.	A10; A11; B10; B11; B16; B23; C10; C11; D10; D11; E10; E11; E16	P	18 (23)	7 (9)
SRO5	Repeats are necessary for calculation of average/confirmation of range.	B32; B33	S	3 (4)	0 (0)
SRO6	Must consider experimental conditions and external factors.	B40; B41; B62; C40; C41; C62; E40; E41	S	8 (11)	3 (4)
SRO7	Measurement is not exact, use approximate value/best estimate.	B14; B20; B21; B22; B61; C21; D20; E20; E21; E23	S	6 (8)	27 (36)
SRO8	Must find best estimate and standard uncertainty.	B24; B50; B60; C24; C50; C51; C52; C53; C61; E50; E51; E52; E60; E61	S	0 (0)	25 (33)
SRO0	Uncodeable responses.	B16; B17; B18; C16; C17; C18; D16; D17; D18; E16; E17; E18; A01; B01; C01; D01; E01; U01	U	14 (18)	0 (0)

Table 4.2 shows that prior to participation in the laboratory course the majority of students (59%) held views rooted in the point paradigm of measurement. At the end of the course, this percentage had been reduced to only 27%, indicating a distinct shift from the point to the set paradigm view. The large percentage of uncodeable pre-intervention responses is due to the number of students suggesting that the distance d should include the length of the ramp, or the height of the table.

The majority of point-reasoners insisted that the true value is attainable given a good enough measuring device with a carefully calibrated scale and suitable experimental apparatus (category SRO3).

Choice A:

“I think that it is exactly 436 mm because the centre of the spot is exactly in line with the 436 mm markings. Therefore it is on the 436 marking.”

(Student 36 - pre)

Choice A:

“They are 10 subdivisions in between 430 and 440 and the mark is on the 6th one.”

(Student 26 - pre)

These statements attest to the claim that the exact value can be determined using an adequately graduated ruler.

Other point-based responses did not question the methods of acquiring the data, but rather suggested that the total value of the distance should include certain dimensions of the ball or block (depending on the experimental context). This view (category SRO4) is exemplified by the following quote taken from students' responses.

Choice C:

“When measuring the distance the ball rolls the circumference (size) of the ball must be taken into account.”

(Student 10 - pre)

Also evidenced amongst responses was the point notion that measurement is exact (category SRO1).

Choice A:

“I’m basing my choice on what information is given and the distance given is 436 mm exactly.”

(Student 86 - pre)

Some students imply that the reading will change on repeating the measurement, only if the experimental conditions are changed, for example, a different starting position will result in a different value of distance, but if the starting point is unchanged, then the final distance will remain the same.

Choice A:

“The ball landed on the spot of paper which when measured was 436 mm exactly so why should it be any different unless the experiment was repeated, it could be different because the exact position of the ball when released could differ.”

(Student 32 - pre)

Adherents to the set paradigm on the other hand, believe that the readings will change with repetition, and that repeats are necessary for calculating averages (category SRO5). Justifications included:

Choice B:

“Because this is not a very reliable test as the paper could have shifted when the ball made contact with it, because the ball does not hit the paper perpendicularly. So this kind of test would have to be performed a few times so that an average ‘ d ’ could be determined.”

(Student 2 - pre)

It also emerged that factors affecting the reading need to be accounted for in order to get a complete value for the measurand (category SRO6).

Choice B:

“When the ball falls down, it is possible that there was an air/wind that push it a little. It is not exactly 436 mm because if you do the experiment again but change the speed of the ball d will not be exactly 436 mm.”

(Student 6 - pre)

Choice B:

“The true value of the distance d can never be known. There is always a level of uncertainty associated with a measurement. Conditions change all the time. Conditions such as the release speed, friction, temp, etc.”

(Student 20 - post)

The set-paradigm idea that the reading offers limited information about the measurand, and thus only a best estimate or good approximation of the value of the distance (categories SRO7 and SRO8), was adopted by 73% of the sample cohort after participation in the course. A typical claim was:

Choice B:

“The value is not certain and there is an uncertainty related to the value which makes the value the best estimate and not the exact value.”

(Student 72 - post)

Students submitted that a complete value for the measurand will include a best estimate and an associated uncertainty (category SRO8). They said:

Choice B:

“We can never measure the exact distance an object has traveled as we need to leave room for uncertainty. When recording distances we need to choose a best estimate and then calculate the standard uncertainty on that result.”

(Student 66 - post)

Choice B:

“The best estimate for the reading is 434.0 mm, as this is only an approximation. The true reading lies between 434.1 and 433.9 (approximately) we can calculate the standard uncertainty. Final answer: best estimate \pm standard uncertainty (65% level of confidence - triangular PDF used).”

(Student 43 - post)

4.2.2 The QD2 and PR1 probes

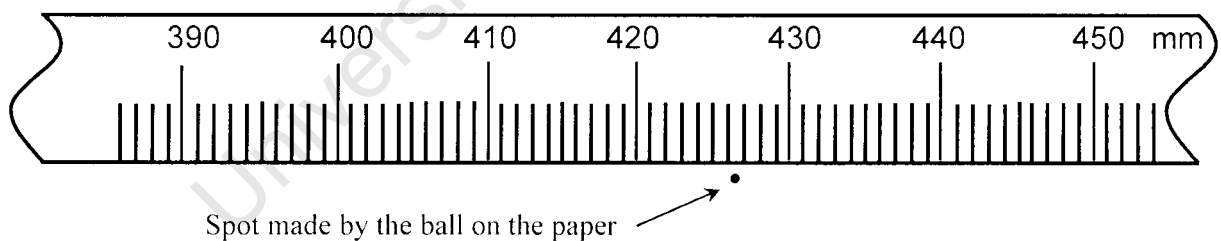
The QD2 and PR1 probes again consider a single reading on an analogue scale, but with the reference point between two graduations.

The QD2 probe:

The given scenario is described below:

The group of students decide to allow the ball to roll again from height $h = 90$ mm.

The student's use the same metre rule to measure the distance d , and what they see is shown below.

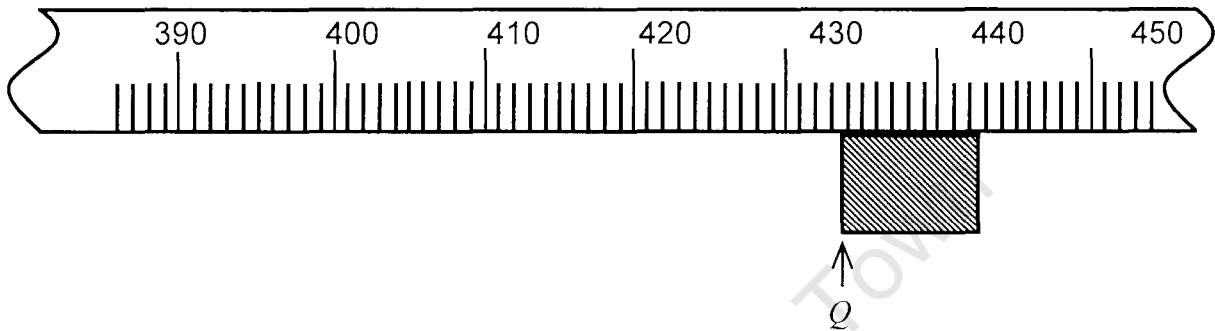


- A: I think that the distance d is exactly 426.5 mm.
- B: I think that the distance d is approximately 426.5 mm.
- C: I think that the distance d is between 426 mm and 427 mm.
- D: I don't agree with any of you.

The PR1 probe:

In the post questionnaire experimental context, the text reads as follows:

One of the groups marks point P and *carefully* lines it up with the zero mark on the metre rule. They then release the block. After the block comes to rest they see that the point Q lines up on the metre stick as shown.



The students then have the following discussion.

- A:** I think the distance d the block has travelled is exactly 433.8 mm.
- B:** I think the distance d the block has travelled is approximately 433.8 mm.
- C:** I think the distance d the block has travelled is between 433 mm and 434 mm.
- D:** I think the distance d the block has travelled is approximately 434.0 mm.
- E:** I don't agree with any of you.

A comparison of responses to the QD2 and PR1 probes completes their analysis.

Table 4.3: *Students' paradigm use in their pre- and post-intervention responses to probes dealing with a single reading on an analogue scale (reference point appears between calibration lines) (n = 76).*

Category	Description	Codes	Paradigm	No. of Students (%)	
				Pre	Post
SRB2	Repeats will give recurring value/accurate answer.	A41; B31	P	1 (1)	0 (0)
SRB3	Better measuring device/more sensitive scale will give accurate/exact results.	A12; B12; B13; B15; C12; C13	P	27 (35)	15 (20)
SRB4	The size, shape, and movement of the ball/block affect the reading.	B10; B11; B23; C10; C11; E11	P	16 (21)	8 (11)
SRB5	Repeats are necessary for calculation of average/best estimate.	B32; B33; D32; E31	S	2 (3)	0 (0)
SRB6	Must consider experimental conditions and external factors.	B40; B41; C40; E40	S	10 (13)	1 (1)
SRB7	Measurement is not exact, use approximate value/best estimate/interval.	B14; B21; B22; B23; B60; C14; C20; C63; D14; E14	S	3 (4)	13 (17)
SRB8	The distance is best represented by an interval given by the best estimate and standard uncertainty.	B51; C21; C52; C54; C62; C63; E50	S	0 (0)	19 (25)
SRB9	The true value of the distance is in a given interval.	B50; C50; C51; C53; C61; E52	S	8 (11)	19 (25)
SRB0	Uncodeable responses.	A01; B01; C01; D01; D60; E01	U	9 (12)	1 (1)

While only 31% of respondents subscribed to the set paradigm of thought in their pre-intervention responses, 68% displayed 'set-thinking' in their post-intervention responses.

The majority of point-based responses claimed that employing finely calibrated measuring devices will yield exact results (category SRB3). Arguments included:

Choice B:

“The ruler is not adequately measured in units less than mm. So it is suffice to say we can’t know for sure if it did indeed travel 426.5 mm because we don’t know for sure if it is a 0.5 mm.”

(Student 67 - pre)

Respondents in the ‘set’ category allude to the fact that measurements are all subject to uncertainties. Typical responses claiming that there is no exact value (category SRB7) were:

Choice B:

“In Physics we can never get an exact distance, only an approximate distance close to the real one.”

(Student 63 - pre)

and

Choice B:

“The distance traveled by the block is approximately 433.8 mm because we do not know the exactly number between 433.0 mm and 434.0 mm. We also cannot say exactly 433.8 mm because there is no exact answer.”

(Student 54 - post)

while responses like

Choice B:

“The distance is not certain because factors such as air friction/resistance are not being taken into account.”

(Student 35 - pre)

and

Choice B:

“I mostly agree with B because you can’t have an exact answer. No scale is small enough to give us one... therefore I think that the reading is approximately 433.8 mm.”

(Student 51 - post)

imply an uncertainty associated with external factors and experimental apparatus, respectively (category SRB6).

The responses

Choice E:

“I agree with student E, because the value of d cannot be estimated by releasing the block once, many readings need to be taken so as to get a good estimate for the value ‘ d ’.”

(Student 87 - post)

and

Choice D:

“one has to repeat the experiment and work out an average distance.”

(Student 32 - pre)

suggest that the average of a few readings is a better approximation of the distance than a single reading (category SRB5).

Among the post-test responses was the concept that the distance is not given by a single value, but is best represented by an interval (categories SRB8 and SRB9). This is evidenced by the following responses:

Choice C:

“The distance d the block has traveled is somewhere between 433.0 and 434.0 mm.”

(Student 62 - post)

Choice E:

“To determine the distance of the block we need to consider the best estimate of the block and the uncertainty of the block. Therefore we can say the block lies within a certain interval.”

(Student 83 - post)

4.2.3 The QD3 and ED1 probes

The QD3 and ED1 probes investigate students' interpretations of a single reading on a digital scale.

The QD3 probe:

The lecturer now comes around with a special meter which has a digital display and uses it to measure the distance d for one of the rolls from $h = 90$ mm. Here is what the electronic meter shows:



The following discussion takes place between the students.

- A: I think the distance d is exactly 423.7 mm.
B: I think the distance d is approximately 423.7 mm.
C: I think the distance d is between 423 mm and 424 mm.
D: I think the distance d is between 423.65 mm and 423.75 mm.
E: I don't agree with any of you

The ED1 probe:

The lecturer now comes around with a special electronic meter which has a digital display and uses it to measure d . Here is what the electronic meter shows:



After the reading has been recorded and the lecturer has left, the following discussion takes place between the students.

- A: Good, we now know that d is exactly 433.0 mm.
B: No, I think that we now know that d is approximately 433.0 mm.
C: I think that d is between 432.5 mm and 433.5 mm.
D: *I think that d is between 431.0 mm and 432.0 mm.*
E: I don't agree with any of you.

Option D was rendered invalid by an undetected typing error. The majority of respondents took note, and limited their choices to the other options. Only three opted for D, but their supporting arguments clearly indicated that the reason was misread, and the responses were coded accordingly.

Table 4.4 compares pre- and post-instruction views on single measurement on a digital scale.

Table 4.4: *Students' paradigm use in their pre- and post-intervention responses to the probes dealing with a single reading on a digital scale (n = 76).*

Category	Description	Codes	Paradigm	No. of Students (%)	
				Pre	Post
SRD1	The electronic meter gives exact/accurate readings.	A12; A15	P	27 (36)	4 (5)
SRD2	Repeats will give recurring value/accurate answer.	A41; E31	P	3 (4)	0 (0)
SRD3	Better measuring device/more sensitive scale will give accurate/exact results.	B13; B15; C13; C15; D13; D15; E15	P	6 (8)	3 (4)
SRD4	The size, shape, movement and mass of the ball/spot/block affect the reading.	A10; B10; B11; D10; D11	P	3 (4)	0 (0)
SRD5	Repeats are necessary for calculation of average/confirmation of range.	B32; B34; C33; E32	S	3 (4)	2 (3)
SRD6	Must consider experimental conditions and external factors.	B40; C16; C40; C41; D41; E40	S	11 (14)	4 (5)
SRD7	Measurement is not exact, use approximate value/best estimate/interval.	B12; B14; B21; B22; B24; C20; C21; E21	S	17 (22)	17 (22)
SRD8	Must find best estimate and standard uncertainty.	C50; C51; E50; E51	S	0 (0)	14 (18)
SRD9	The true value can never be known, associated uncertainty not zero.	B60; B62; B63; C62; C63; E61; E62; E63	S	0 (0)	30 (39)
SRD0	Uncodeable responses.	A01; B01; C01; D01; E01; U01	U	6 (8)	2 (3)

Table 4.4 shows that the emerging trend from the pre- and post- comparisons of probes dealing with single readings continues in the comparison of how students' deal with a single digital reading. Even more prominent pre-post paradigm shifts occur, with point paradigm-based reasoning decreasing from 52% to 8%, and set-reasoning increasing from 40% to 87%.

Compared to the findings for the single analogue reading (see Tables 4.2 and 4.3), the findings for the single digital reading show some remarkable differences in the distribution of the responses over the different categories. The slightly higher initial set reasoning for the digital reading is almost entirely due to the large percentage of responses stating that the digital reading is not exact (22% in SRD7) with much lower percentages for the analogue reading (8% and 4% for SRO7 and SRB7, respectively). For the analogue reading, initial point reasoning is mainly expressed through a belief that a finer scale graduation will provide the true value (see categories SRO3 and SRB3). In contrast, point reasoning for the digital reading is expressed in acceptance that the reading is the exact (true) value because of the digital scale being used (see category SRD1 below). After the course, the faith in the exactness of the digital scale has all but disappeared, but the point notion that a finer scale will allow identifying the true value of the measurand still has some currency.

As suggested above, most 'point-thinkers' placed absolute faith in the electronic meter. They asserted that an electronic meter always gives exact results (category SRD1).

Choice A:

"What we see on the electronic meter is what we get."

(Student 82 - post)

Choice A:

"I think the distance d is exactly 423.7 because even the electronic meter got the same distance."

(Student 36 - pre)

Choice A:

“Electronic things always make no mistakes.”

(Student 105 - pre)

In contrast, most of the respondents claimed that not even the use of an electronic meter can eliminate all uncertainties from a measurement. The response

Choice B:

“We don’t exactly know what the value of d is because even a digital display by the electronic meter is not made to show all the numbers or the value therefore it is an approximation of the distance.”

(Student 54 - post)

suggests that there is uncertainty associated with the calibration of the meter (category SRD7), and the argument

Choice E:

“Even an electronic meter cannot say accurately what in each case the result will be, it only sorts out the case so you are accurate when measuring the distance the ball landed each time, but external factors, and the position the ball was released could still affect that distance d .”

(Student 32 - pre)

states that the measurement is affected by external factors and the experimental setup (category SRD6).

They propose that the uncertainties should be dealt with by repeating the measurements and calculating the average (category SRD5),

Choice E:

“The lecturer should also use the electronic meter several times and then calculate an average, he can’t just do it once and say that that is a final answer.”

(Student 69 - pre)

or by finding the best estimate and standard uncertainty (category SRD8)

Choice C:

“Because when using a digital instrument we can say that the uncertainty of a value is $u(d) = 0.29$ and therefore $d = 433.0 \pm 0.29$ which does fall within 432.5 and 433.5. we have defined an interval in which the true value lies as 433.0 is our best estimate.”

(Student 110 - post)

A common view expressed by 39% of post-test respondents is that the true value of a measurand can never be known (category SRD9). This is attested to by the following response:

Choice B:

“Because even if an electronic meter is used it will never give the exact value, in fact no apparatus can ever give the exact value because a ‘true value’ of the measurand can never be known. That is why the student is approximating the distance d .”

(Student 102 - post)

4.3 Data processing probes: Students’ ideas about the average of a set of measurements

4.3.1 The UR probes

The UR probes explore how students relate a given collection of repeated readings to the value of the measurand.

The UR probe:

A third group of students releases the block five times from point P. The five values they obtain for d are shown below.

<u>Release</u>	<u>d (mm)</u>
1	436
2	426
3	434
4	430
5	434

The students then discuss what to write down for their final result for d .

Table 4.5 summarizes the students' thoughts on what the measurand will be.

Table 4.5: *Students' paradigm use in their pre- and post-intervention responses to the UR probe ($n = 76$)*

Category	Description	Codes	Paradigm	No. of students (%)	
				Pre	Post
UR1	The average represents the best estimate/best approximation of the distance.	10; 11; 20	S	34 (45)	43 (57)
UR2	The interval given by average and uncertainty is the best representation of the value of the distance.	40; 41	S	7 (9)	30 (39)
UR3	The highest/lowest/middle/recurring data value is the value of d .	12; 30; 50; 60	P	29 (38)	3 (4)
UR0	Uncodeable responses.	U00	U	6 (8)	0 (0)

The pre-intervention probe saw 38% of respondents subscribe to the point school of thought when relating a collection of data values to the measurand; this percentage was reduced to just 4% at the end of the course. Students appeared to have acquired a firm

grasp of the ideas of the average value as the best estimate of the measurand, and of including standard uncertainty in the representation of the measurand.

4.3.2 The UA probes

The UA probes are designed to investigate students' understanding of the average of a data set. Both the UA1 probe and the UA2 probe formed part of the pre and post questionnaires. The core ideas gleaned from the probe responses are thus presented as pre and post comparisons in both cases.

The UA1 probe:

The UA1 probes purpose to determine students' perceptions of the relationship between the average and the measurand. The probe text is here reproduced.

The students decide to calculate the average of their readings for d , which is 432.0 mm.

<u>Release</u>	<u>d (mm)</u>
1	434.5
2	432.0
3	435.8
4	426.6
5	431.1
Average	432.0

They then discuss what the average for the distance d tells them.

- A:** I think that the distance d is exactly 432.0 mm.
B: I think that the distance d is approximately 432.0 mm.
C: I think that the distance d is somewhere between 431.5 mm and 432.5 mm.
D: I think that the distance d is somewhere between 426.6 mm and 435.8 mm.
E: I don't agree with any of you.

Table 4.6: Students' paradigm use in their pre- and post-intervention responses to the UA1 probe ($n = 76$).

Category	Description	Codes	Paradigm	No. of students (%)	
				Pre	Post
UA11	The average is the true value.	A10	P	1 (1)	0 (0)
UA12	The average is exact and remains unchanged after repeats.	A30; A31; A32; E32	P	5 (7)	2 (3)
UA13	The readings are not the same so the average is the true value.	A41	P	1 (1)	0 (0)
UA14	The average is close to the true value/best estimate/approximation.	B10; B11; B12; B13; B16; C11; C13; C14; C15; C16; D11; D14; E13; E14; E16	S	25 (33)	24 (32)
UA15	The average is not exact and may/will change after repeats.	B30; B31; B33; C30; C33; D30; D31; D33; D34; E30; E31; E33	S	15 (20)	14 (18)
UA16	Error must be accounted for.	B20; B21; C20; D20	S	5 (7)	10 (13)
UA17	The readings are not the same so the true value can't be known/is in same interval as readings.	B41; B42; B44; D40; D41; D44; E42	S	20 (26)	5 (7)
UA18	The standard uncertainty must be calculated to get the interval in which the true value lies.	B71; C70; D72; E70; E71	S	0 (0)	21 (28)
UA19	The average and the true value are unrelated.	E60	U	1 (1)	0 (0)
UA10	Uncodeable responses.	U00; A01; B01; C01; D01; E01	U	3 (4)	0 (0)

The UA1 probe does not differentiate well for point and set reasoning. Table 4.6 shows that the point perspective is held by only 9% of respondents prior to the laboratory course, and by 3% upon completion of the course. The changes in students' understanding of the meaning of an average occur within the set paradigm. Views shift from the notion that since the readings are not identical, the true value cannot be selected but lies within the

interval given by the readings (category UA17), towards the idea that a standard uncertainty must be calculated to determine the interval within which the true value lies (category UA18).

Point reasoning is evident in the following responses:

Choice A:

“The average in actual fact is the accurate distance.”

(Student 86 - pre)

Choice A:

“The average was found after many experiments, so even if the experiments can be performed a dozen times, it will be found that the average is still 432mm.”

(Student 49 - pre)

Choice A:

“If something gives you different figures, and you are looking for the exact point, you have to use the average.”

(Student 58 - pre)

These quotes, reproduced from students' responses, epitomize their distinctly point perspective. The underlying rationale includes the idea that varying readings gives one no choice but to accept that the average is the true value (category UA13), that the average is exact or accurate, and will remain unchanged after further repetition of the experiment (category UA12), and the unambiguous assertion that the average is the true value (category UA11).

An alternative logic which produces the idea that the true value cannot be conclusively decided on because more readings will change the average (category UA15), was also evident among responses.

Choice E:

“The average of d won't stay the same if another roll was done, for example, if there were 7 rolls done, the average will maybe be 436 mm...”

(Student 23 - pre)

The 86% majority of set-associated responses identified in the pre-test responses increased to 97% in the post-test analysis. In both cases, a third of respondents claimed that the true distance is in a small interval around the average, or that the average is the most likely value, or best approximation, of the true distance (category UA14). Supporting statements included:

Choice C:

“It, (the distance d), is closest to the average and has a 0.5 allowance.”

(Student 56 - pre)

Choice B:

“The average value is not the exact value it is simply an approximation of what the exact value might be.”

(Student 64 - post)

Choice B:

“Because that is the best estimate for d and it cannot be exactly 432.0 mm as it differs, so that is why it is just an approximation.”

(Student 88 - post)

and

Choice B:

“It's the closest we can get to the exact distance.”

(Student 82 - pre)

Appearing in both pre and post responses was the concept of representing the value of the measurand by an interval. Justifications included the opinion that the true distance is likely to be in an interval defined by the range of the readings (category UA17),

Choice D:

“Because all the answers they got were ranging between 426 and 436, so it [d] should be somewhere between that.”

(Student 27 - pre)

Choice D:

“Each result obtained was between 426 mm and 436 mm, showing that the distance could very well be anywhere between 426mm and 436 mm.”

(Student 68 - pre)

and the decidedly set notion that the true value is given by the best estimate and its standard uncertainty (category UA18).

Choice C:

“The exact value will be $432.0 \pm u(d)$ if standard uncertainty is calculated which means that it lies somewhere between 431.5 and 432.5 mm.”

(Student 89 - post)

The UA2 probe:

The UA2 probes question whether or not the average of a given set of readings gives any indication of what the next reading will be.

The students have 5 readings for d :

<u>Release</u>	<u>d (mm)</u>
1	434.5
2	432.0
3	435.8
4	426.6
5	431.1
Average	432.0

The students now discuss what reading they will get for d if they repeat the experiment once more (a sixth time).

- A:** I think we will get a reading of 432 mm.
- B:** I think we will get a reading somewhere between 431 mm and 433 mm.
- C:** I think we will get a reading somewhere between 426 mm and 436 mm.
- D:** I think that the new reading can have any value.
- E:** I don't agree with any of you.

Table 4.7: *Students' paradigm use in their pre- and post-intervention responses to the UA1 probe (n = 76).*

Category	Description	Codes	Paradigm	No. of students (%)	
				Pre	Post
UA21	The next reading will be such that the average remains unchanged.	A30; E30	P	3 (4)	1 (1)
UA22	The next reading can't be predicted because the previous readings have no discernible pattern.	C11; C12; C14; D11; D12; D13; D14; E10; E12; E14; E15	S	17 (22)	18 (24)
UA23	The next reading will be affected by experimental conditions and external factors.	C20; C21; D20; D21; E20; E21	S	14 (18)	30 (39)
UA24	The next reading will be in the range of/close to the previous readings.	B16; C10; C50	S	25 (33)	8 (11)
UA25	The next reading will be in the same interval as/close to the average.	B31; B32; B40; B41; C31; D40; E31; E40; E42	S	9 (12)	8 (10)
UA26	The next reading can't be predicted because it will change the average.	C41; D41; E41	S	3 (4)	7 (9)
UA27	The next reading will be in the interval given by the best estimate/average \pm standard uncertainty.	C60; C61; C70; E60; E61	S	0 (0)	4 (5)
UA20	Uncodeable responses.	U00; A01; B01; C01; D01; E01	U	5 (7)	0 (0)

The responses to the UA2 probes were predominately set-based, with 89% of the sample group before instruction, and 99% after instruction, subscribing to that particular

paradigm. As with the UA1 probe, this probe differentiates poorly for point and set reasoning, and the changes in students' understanding occur within the set paradigm. The majority-held views, both before and after instruction, are that the next reading will be affected by external factors and experimental conditions (category UA23), and that the next reading will be in the range of previous readings (category UA24).

The few point responses proposed that the next reading should be such that the average remains unchanged (category UA21), as illustrated below.

Choice E:

"In order for the average to remain 432 mm, the reading for the sixth time should be 432 mm exactly."

(Student 86 - pre)

The rest of the respondents are divided into those who believe that it is impossible to make any predictions about the next reading, and those who expect the next reading to be in some way related to previous results.

The former group is further divided by the different reasons behind their belief. Some claim that there is no discernible pattern in previous readings (category UA22), on which to base a prediction,

Choice D:

"I can't seem to find a pattern in the readings and therefore it can be any value."

(Student 11 - pre)

others say that attempts to predicate the next reading are pointless in light of the uncontrollable external factors and experimental conditions which are likely to affect the reading (category UA23).

Choice D:

"Every time we roll the ball the external factors change. The wind might be blowing harder so we can't predict what will happen next."

(Student 20 - pre)

Choice D:

“there is no mention of conditions around the experiment being maintained the same. To be able to predict you have to keep everything from the first to the last roll the same.”

(Student 30 - pre)

A final subdivision concludes that the next reading will change the average and can thus not be pre-determined (category UA26).

Choice D:

“It is a new roll altogether which will ultimately bring a new average.”

(Student 22 - pre)

Proponents of the idea that the next reading will be related to previous results, argue that the next reading will fall in the same range as previous readings (category UA24),

Choice C:

“After 5 times of rolling the ball the reading has stayed between 426 and 436 mm. If the ball is rolled for the sixth time the reading will remain between 426 and 436 mm.”

(Student 19 - pre)

that the next reading will be close to the average (category UA25),

Choice C:

“Seeing that the average value is 432 mm, when the ball is rolled again, the reading will either be a little more or a little less than 432 mm.”

(Student 112 - pre)

and that the next reading will be in an interval defined by the best estimate of the distance and its standard uncertainty (category UA27).

Choice C:

“As you can see from the previous question, the students obtain for d to be, $d = 432.0 \pm 1.59$; 68% confidence level. That implies that the next value for d must lie between 426.6 and 435.8 mm, because they are 68% confident that it will.”

(Student 23 - post)

4.4 Data comparison probes: Students' ideas about comparing data sets and averages

4.4.1 The SMDS and DMSS probes

The SMDS and DMSS probes deal with the comparison of data sets. The students are presented with two data sets and their averages and asked to decide whether or not they give equally valid results.

The SMDS (Same Mean Different Spread) probe:

Two groups of students, both of whom had decided to release the block 5 times from point P , compare their measurements for d . The values for the five releases are shown below, together with their averages.

<u>Release</u>	<u>A: d (mm)</u>	<u>B: d (mm)</u>
1	443.5	446.8
2	432.8	459.4
3	424.4	410.5
4	439.6	423.3
5	434.9	435.0
Average	435.0	435.0

A: Our measurement for d is better than yours.

B: Our measurement for d is just as good as yours.

C: I don't agree with either of you.

Table 4.8: *Students' paradigm use in their pre- and post-intervention responses to the SMDS probe (n = 76).*

Category	Description	Codes	Paradigm	No. of students (%)	
				Pre	Post
SMDS1	The results are equally good since the averages are identical.	B20; B21; B22; B23; B26; B27; B29; B30; B61; B64; B65; B70; C50; C60	P	54 (71)	26 (34)
SMDS2	The spread of data values must be compared.	A10; A11; A12; A13; A14; A15; A16; A62; A63; A64; B10; B11; B12; C10; C12; C66; C91	S	16 (21)	36 (47)
SMDS3	The average and the standard uncertainty are important.	A20; A21; A22; A25; B24; B81; C80	S	2 (3)	13 (17)
SMDS0	Uncodeable responses.	A01; B01; B60; C01; U01	U	4 (5)	1 (1)

Before the course, there was a significant 1 to 3 split in terms of set and point reasoning respectively, with only 18 of the 76 respondents (24%) considering the range of the data values and the standard uncertainty relevant in the comparison of the data sets, and 71% claiming that the results were equally good since the averages were identical. In the post instruction responses the imbalance shifted in favour of the set paradigm with 64% of respondents realizing the importance of including the data spread and standard uncertainty in the comparison of results, and only 34% maintaining that it was sufficient to compare averages.

The DMSS (Different Mean Similar Spread) probe:

In the DMSS probe, the averages are not the same:

Two groups of students compare their measurements for d . The values for the five releases are shown below, together with their averages.

<u>Release</u>	<u>A: d (mm)</u>	<u>B: d (mm)</u>
1	439.5	435.6
2	438.4	439.2
3	433.1	428.0
4	422.8	433.1
5	431.3	438.3
Average	433.0	434.8

- A:** Our result for d agrees with yours.
B: No, our results do not agree.
C: I don't agree with either of you.

Analysis of responses to the DMSS probe showed similar results. As can be seen in Table 4.9, 85% of students expressed pre-intervention views grounded in the point paradigm, with post-intervention responses indicating a significant shift to the set paradigm. The number of respondents who asserted that the comparison of the data set is enabled by the examination of the uncertainties about the averages and the spread of the data values, increased from 6 (8%) to 57 (75%).

Table 4.9: Students' paradigm use in their pre- and post-intervention responses to the DMSS probe ($n = 76$).

Category	Description	Codes	Paradigm	No. of students (%)	
				Pre	Post
DMSS1	It depends on how close the averages are.	A20; A21; A22; A24; A26; A50; B20; B21; B22; B24; B25; B26; B50	P	53 (70)	10 (13)
DMSS2	It depends on the degree of correspondence between individual data values in the two sets.	A10; A40; A41; B10; B31; B40; C40	P	11 (14)	7 (9)
DMSS3	It depends on what the true value is.	B60; B61; C60	P	1 (1)	1 (1)
DMSS4	It depends on the relative spreads of the data.	A12; A13; B12; B32	S	4 (5)	10 (13)
DMSS5	It depends on both the averages and the uncertainties.	A30; A31; B30; C20; C30; C70; C71	S	2 (3)	47 (62)
DMSS0	Uncodeable responses.	A01; B01; B80; C01; U01	U	5 (7)	1 (1)

4.5 Understanding uncertainty probes: Students' ideas about uncertainty in measurement

4.5.1 The NU1 and PX1 probes

The NU1 and PX1 probes deal with students' beliefs about whether or not uncertainties associated with measurement can be reduced to zero.

The NU1 (No Uncertainty 1) probe:

The NU1 probe asks whether improving experimental skill can lead to finding the true or exact value of the measurand.

The context and discussion are presented below:

When they are finished, the two groups discuss how they can improve their rolling ball experiment next time.

- A:** If we practice enough, all our readings will be the same. Then we will know the true value of d .
- B:** No, even if all your readings are the same, you will still not know the true value of d .

The views expressed by the respondents were reduced to six main categories.

Table 4.10: *Students' paradigm use in their responses to the pre-intervention NU1 probe ($n = 100$).*

Category	Description	Codes	Paradigm	No. of students (%)
NU11	Practice and careful work yields true value.	A10; A11; A12; A13	P	28 (28)
NU12	Same readings, recurring value is true value.	A20	P	7 (7)
NU13	Experimental conditions change, can't control external factors, so can't know true value.	B11; B12; B13	S	35 (35)
NU14	The readings will always vary, so can't know true value.	B21	S	10 (10)
NU15	No true value, so approximate.	B30	S	6 (6)
NU16	No true value, so use average.	B31	S	10 (10)
NU10	Uncodeable responses.	U01; A01; B01; N00	U	4 (4)

Of the 100 respondents to the NU1 probe, 35 believe that it is possible to find a value of a measurand with no associated uncertainty. This is one of the defining characteristics of the point paradigm. The majority of these respondents claim that this true or exact value can be found through practicing experimental techniques and paying careful attention to the measurement process (category NU11). Supporting arguments include:

Choice A:

“Practice makes perfect, so if the same experiment is done very carefully a perfect reading will be obtained.”

(Student 8 - pre)

Choice A:

“If they are careful with the height and use good apparatus and a flat floor surface, they will have the same readings at all times.”

(Student 64 - pre)

Choice A:

“If a proper method is used then accurate results would be obtained. If the height remains the same and there are no external factors influencing the distance reached, then the distances obtained should be equal to one another.”

(Student 66 - pre)

and

Choice A:

“If the conditions are similar, no matter how many times we do the experiment, the readings will always be the same.”

(Student 38 - pre)

It is asserted, by 7% of respondents, that the true value is the one that recurs when measurements are repeated (category NU12). They argue:

Choice A:

“If all the readings are the same continuously, then the true value of d is that reading.”

(Student 46 - pre)

Choice A:

“Because if we have the same readings, this means that the true value of d is that reading.”

(Student 111 - pre)

Fundamental to the set paradigm is the notion that all measurements are subject to non-reducible uncertainties. Most of the responses to the NU1 probe support this idea.

The idea that the uncertainties can be attributed to uncontrollable experimental and external factors (category NU13) is held by 35% of respondents. Among the responses were:

Choice B:

“They could not find the true value of d because the reading on the metre stick will always vary.”

(Student 7 - pre)

and

Choice B:

“We will never know the true value of d . It doesn't matter how much a human practices the experiment, there are other things in nature that affect it, such as air resistance and a difference in speed due to gravity...”

(Student 45 - pre)

Another reason cited for the true value being unattainable was the variations in readings when measurements are repeated (category NU14).

Choice B:

“One would still not know the true value of d because we will be getting different values after each roll.”

(Student 12 - pre)

Choice B:

“The values of d vary too much to obtain a ‘true’ answer for it.”

(Student 44 - pre)

A somewhat more advanced understanding is evidenced by the claim, of 16 respondents, that there is no true value, and we can, at best, approximate the distance (category NU15) or use the calculated average (category NU16). Typical responses were:

Choice B:

“The ‘exact’ d will never be found. Because every time you roll it, you will hardly ever come up with the same result. ...therefore we can only assume an average or an approximate value, but never the exact value of d .”

(Student 74 - pre)

Choice B:

“there is no true reading for d but they can only give the approximate.”

(Student 18 - pre)

Choice B:

“I think we will never know the true value of d . We will have to work with the average.”

(Student 3 - pre)

and

Choice B:

“I believe that practice makes perfect but you calculate an average not an exact distance. It’s the closest you can get to the true value of d .”

(Student 82 - pre)

The PX1 probe:

The PX1 probe asks a similar question to the one posed by the NU1 probe, only in this case the possibility of finding the true or exact value through sophisticated experimental design, rather than experimental skill, is explored.

The students continue to discuss the experiment.

A: Isn’t it sad that nobody can ever know the real value of d .

B: That’s not true! If we had the money we could design an experiment which would give us the real value of d .

Five main ideas emerged from the responses to the PX1 probe. These are listed in Table 4.11.

Table 4.11: *Students' paradigm use in their responses to the post-intervention PX1 probe (n = 117).*

Category	Description	Codes	Paradigm	No. of students (%)
PX11	Given enough money, it is impossible to create a measuring device good/sensitive enough to determine the true value.	B10; B11	P	7 (6)
PX12	The true value can never be known, only best estimate/most likely interval.	A20; A21; A22	S	31 (26)
PX13	There are always uncertainties associated with experimental conditions and external factors.	A30; A31; A32; A33	S	41 (35)
PX14	It is impossible to create a measuring device good/sensitive enough to determine the true value.	A40	S	35 (30)
PX15	All measurements are approximate.	A10	S	2 (2)
PX10	Uncodeable responses	U01; A01; B01	U	1 (1)

A mere 6% of the 117 students responding to the post-intervention PX1 probe maintained that it is possible to reduce uncertainty to zero and thus get the true value of a measurand. They asserted that this can be achieved by creating a sufficiently sensitive measuring device, and acquiring the skills to use it (category PX11).

Choice B:

“The value of d can be found with the help of an instrument that is very sensitive. For this lots of money is needed. The experiment and instruments are not sensitive enough and that is why standard uncertainties have to be considered.”

(Student 11 - post)

Choice B:

“There are some machines that can be used to give better results but there are some conditions that you have to know how to use the instrument perfectly because the standard uncertainty are caused by lack of knowledge of using some measurand.”

(Student 25 - post)

In contrast, of the responses which were consistent with the set paradigm (109 of 117), 32% expressed the belief that it is impossible to create a device sensitive enough to determine the true value (category PX14).

Choice A:

“No instrument can be infinitely correct. The exact value can never be known.”

(Student 43 - post)

Choice A:

“All the money in the world couldn't allow to make an instrument that could measure the exact value. An analogue reading would always have some human error possibility and have an electronic meter with an infinite no. of decimal places is not possible, there isn't enough place in the universe for something like that.”

(Student 81 - post)

Experimental conditions and external factors appear as usual contributors to the uncertainties associated with measurements (PX13). Their role is explained as follows:

Choice A:

“In every experiment, and every measurement there is always possibilities of uncertainties. Even if you have all the money in the world there is no possible way to rule out every possible factor that can cause uncertainties. For one you will not be able to manufacture machinery without flaws and there is also human error to consider.”

(Student 57 - post)

Choice A:

“Because of uncertainties which affects out experiments like zero reading, internal calibration, pressure influence and so on.”

(Student 28 - post)

While some response were simply statements that all measurements are approximate (category PX15),

Choice A:

“All measurements are approximations. There will always be a better scale to give us a more correct value but never the exact.”

(Student 63 - post)

others suggested that since the true value can never be known, it should be represented by an interval given by the best estimate and standard uncertainty (category PX12).

Choice A:

“Because it is true that the exact value for d , or any measurand will never be known. Which is why one should report the final result of an experiment with a standard uncertainty and a level of confidence, using probability density functions (pdf's).”

(Student 102 - post)

Choice A:

“The real value of d can never be known but only the approximate of d can be known. We use the triangular pdf to find the best estimate of the value but not the exact.”

(Student 18 - post)

Careful consideration of the responses to the NU1 and PX1 probes reveal that the larger proportion of the sample cohort (61% in NU1 and 93% in PX1) appreciates the fact that uncertainties cannot be reduced to zero, and quantities being measured are thus never exact.

4.6 Nature of measurement probes: Students' ideas about the nature of measurement

The AE1 probe:

The AE1 probe investigates students' ideas about the nature of measurement. The text of the probe is reproduced below.

The following discussion takes place between some of the students.

- A: It is ok to have approximate measurements for everyday use but for physics, the measurements have to be exact.
- B: It all depends on whether you want to find a mathematical formula or not.
- C: No, all measurements are always approximate.
- D: No, all measurements are always exact.
- E: I don't agree with any of you.

Option D was misread as “**Not** all measurements are always exact.” The responses were coded as C or E depending on the reasons cited for the choice.

The main ideas revealed by the responses are tabulated in Table 4.12 below.

Table 4.12: *Students' paradigm use in their responses to the pre-intervention AE1 probe (n = 100).*

Category	Description	Codes	Paradigm	No. of students (%)
AE11	In Physics/Science, societal consequences of inaccuracy are greater.	A11	P	9 (9)
AE12	Physics/Science must be accurate/have exact values.	A12; A13; A14; A20; E14; E20	P	17 (17)
AE13	Exact measurements are required to find the true value/correct answer.	A30; E30	P	5 (5)
AE14	The purpose of the measurement determines whether exact or approximate values are required.	B10; B11; B12; E40	P	14 (14)
AE15	In Physics/Science cannot get exact values, only approximate ones.	C10	S	3 (3)
AE16	Measurements are always approximate.	C40; C41	S	14 (14)
AE17	Experimental and external factors affect measurements.	C51; C52; C53; E51	S	21 (21)
AE10	Uncodeable responses.	U01; A01; B01; C01; D01; E01	U	17 (17)

When presented with the task of deciding when measurements should necessarily be exact and when approximations will suffice, 45% of respondents fell in with the characteristic point idea that exact measurements with no uncertainty are not only possible but often required.

While some students asserted without qualification that physics requires exact measurements because physics, or science, must be accurate (category AE12),

Choice A:

“Tolerances are factored into everyday measurements... but purest physics has to be absolutely accurate, otherwise what’s the point?”,

(Student 62 - pre)

others cited various reasons including consequences to society, requirements for calculations and research, and just getting the right answer (categories AE11 and AE13). Among these arguments were:

Choice A:

“It is important to have exact measurements in physics because physics people work with electricity and computers and other things so if the measurements are not exact it could be dangerous.”

(Student 64 - pre)

Choice A:

“Having exact measurements in physics helps when dealing with calculations.”

(Student 72 - pre)

Choice A:

“Whenever you are doing an experiment or measuring the readings must be exact to receive the best result.”

(Student 15 - pre)

Choice A:

“If you are creating an experiment and want to base research on the information taken from the experiment then measurements should be exact.”

(Student 16 - pre)

Choice A:

“So that you can be able to get the exact, real answers, therefore in physics the measurements have to be exact.”

(Student 27 - pre)

Choice A:

“Measurements have to be accurately done in order to obtain the correct answer.”

(Student 40 - pre)

Measurements can be exact or approximate depending on their purpose (category AE14). This claim is supported by 14% of respondents. Again, the point idea that a measurement can be without error and in fact, be the true value, was evident.

Choice B:

“I think it depends on what you are trying to find. Example: when you are working with chemicals you have to be exact to find the results you are looking for.”

(Student 3 - pre)

Choice B:

“Approximate measurements are often used, but when trying to find a mathematical formula, measurements need to be exact.”

(Student 81 - pre)

Two respondents in this category suggested, rather interestingly, that the average can be used in cases where the experiment is unimportant, since the average itself is useful only when an exact measurement can't be made.

Choice B:

“If the experiment is of no use or importance, an average is fine. Since averages is relatively useless, it has no importance.”

(Student 47 - pre)

Choice B:

“It does not necessarily mean measurements will always be approximate. Therefore using average measurements might do the trick.”

(Student 71 - pre)

The latter response shows a misunderstanding of the word ‘approximate’. It appears that it was intended to mean ‘accurate’ or ‘exact’. Both responses betray a dire lack of understanding of what an average is.

Of the 100 responses, 38 are associated with the set paradigm. They claim that all measurements are subject to uncertainties, and consequently not exact. Of these, 17 asserted that there are no exact values in physics or science, that measurements are always

approximate, and that one always approximates in experiments (categories AE15 and AE16). The justifications include:

Choice C:

“In physics you can only get a distance that is close to the real one. Never exact.”

(Student 63 - pre)

Choice C:

“One can never be sure of a result received, therefore most or all results are usually approximate.”

(Student 68 - pre)

Choice C:

“It is very rare to find or to have exact measurements as experiments’ figures are approximated.”

(Student 14 - pre)

Choice C:

“It is impossible to find an exact measurement. The distance the ball rolls could be different every time. All measurements are therefore approximate.”

(Student 17 - pre)

Others suggested that it is experimental and external factors which render all measurements approximate (category AE17).

Choice C:

“I don’t think we will ever get exact answers but using the most correct tools will help us to get the closest answers possible to the true answer.”

(Student 13 - pre)

Choice C:

“Because we all use different apparatus when performing our experiment.”

(Student 77 - pre)

Choice C:

“No measurement is exact as a result of external factors.”

(Student 101 - pre)

Choice C:

“You cannot always have an exact measurement. External factors will always be there to make your measurements approximate, even in physics.”

(Student 78 - pre)

The responses to this probe suggest that the widely held view among the students in this cohort is that measurements can be either exact or approximate depending on their purpose, and that scientific disciplines including physics and mathematics require exact measurements.

University of Cape Town

Chapter 5: Findings - Cross Probe Analysis

In this chapter, comparisons of the paradigms underlying students' responses before and after participation in the laboratory course are presented.

The tables show the pre-intervention to post-intervention paradigm shifts for individual students, and the histograms illustrate the pre-post change in the distribution of percentages (and numbers) of the total cohort of students in each paradigm category. The analysis process is described in Section 3.6.4.

These results highlight the level of success of the laboratory course in achieving its teaching objectives in individual aspects of measurement, and overall.

5.1 Understanding of measurement in data processing

The data processing probes include three questions relating to single readings and three questions concerned with multiple readings. Students were classified into either the point paradigm, set paradigm, mixed paradigm, or not classified category based on their responses to all three probes in each area.

The combination of three set-based responses (SSS) and that of two set-based responses and one unclassifiable response (SSU) relegated students to the set paradigm category; point-thinkers were similarly identified and put in the point paradigm category; a combination of point and set responses (PPS, PSS and PSU) saw students assigned to the mixed paradigm category; and the remaining response combinations (PUU, SUU and UUU) were categorised as 'not classified'.

5.1.1 Single measurements

The data processing probes focusing on single measurements are QD1, QD2, QD3, PR1, PR2 and ED1.

Table 5.1 compares students' paradigm use in dealing with single readings before and after instruction.

Table 5.1: *Pre- and post-instruction paradigm distributions among students when processing a single reading (n = 76).*

		Post – instruction (PR1, PR2, ED1)			Total
		Consistent Point Paradigm	Mixed Paradigm	Consistent Set Paradigm	
Pre - instruction (QD1, QD2, QD3)	Consistent Point Paradigm	2 (3%)	15 (20%)	12 (16%)	29 (38%)
	Mixed Paradigm	0 (0%)	18 (24%)	17 (22%)	35 (46%)
	Consistent Set Paradigm	0 (0%)	4 (5%)	4 (5%)	8 (11%)
	Not Classified	2 (3%)	0 (0%)	2 (3%)	4 (5%)
Total		4 (5%)	37 (49%)	35 (46%)	76 (100%)

Only 2 of the 76 students surveyed (3%) are consistently identified as point-reasoners before and after the laboratory course, and 16% of students initially classified as point-thinkers shifted to set-based reasoning in their post-instruction responses. A quarter of the sample cohort (25%) went from using mixed paradigms (22%) and unclassifiable paradigms (3%), prior to instruction, to the set paradigm after instruction.

The overall pre- to post-course paradigm distributions for processing single readings is illustrated in Figure 5.1.

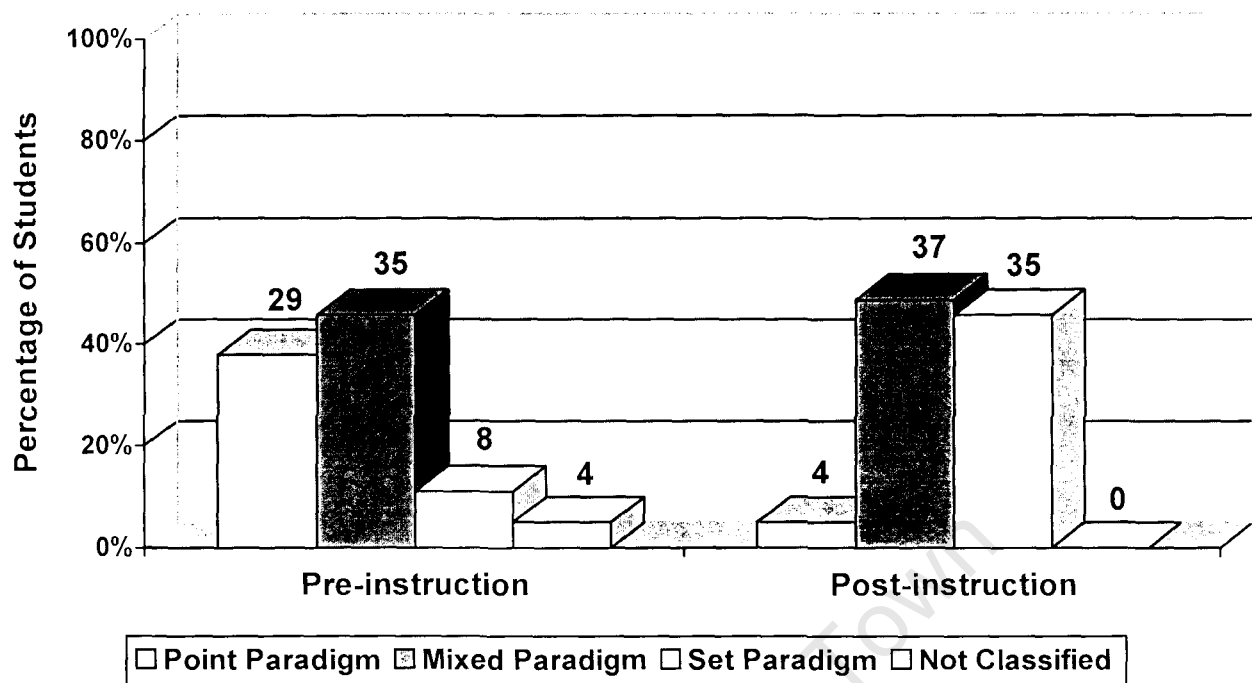


Figure 5.1: Comparison of paradigm distributions for probes relating to a single reading before and after instruction.

Subsequent to participation in the laboratory course, almost half of the respondents were in the mixed paradigm category (49%); the number of set-reasoners had increased from 8 (11%) to 35 (46%), and only 5% of respondents still held consistently point-based views.

5.1.2 Multiple measurements

The data processing probes focusing on the average of a set of measurements are UR, UA1 and UA2.

The pre-post shift in students' use of paradigms when dealing with data sets is detailed in Table 5.2.

Table 5.2: *Pre- and post-instruction paradigm distributions among students when processing multiple readings (n = 76).*

		Post - instruction (UR, UA1, UA2)		Total
		Mixed Paradigm	Consistent Set Paradigm	
Pre - instruction (UR, UA1, UA2)	Consistent Point Paradigm	0 (0%)	2 (3%)	2 (3%)
	Mixed Paradigm	4 (5%)	31 (41%)	35 (46%)
	Consistent Set Paradigm	2 (3%)	34 (45%)	36 (47%)
	Not Classified	1 (1%)	2 (3%)	3 (4%)
Total		7 (9%)	69 (91%)	76 (100%)

The idea of using the average as the best representation of a set of measurements is widely taught as part of the mathematics curriculum at secondary schools in South Africa. It is therefore, unsurprising that close to half of the students (47%) demonstrated a grasp of this concept by employing the set paradigm prior to the start of the university laboratory course. By the end of the course, the 3% of students originally holding point-based views, and 31 of the 35 students (41%) originally employing mixed paradigms, had also shifted to the set paradigm.

These findings show a marked contrast to students' views of processing single measurements, as presented in section 5.1.1. For single measurements, only 11% of students consistently used the set paradigm before instruction. Although the number of set-thinkers did increase after instruction, they still represented less than half of the sample cohort (46%).

The overall paradigm shifts for repeated measurements are illustrated in Figure 5.2.

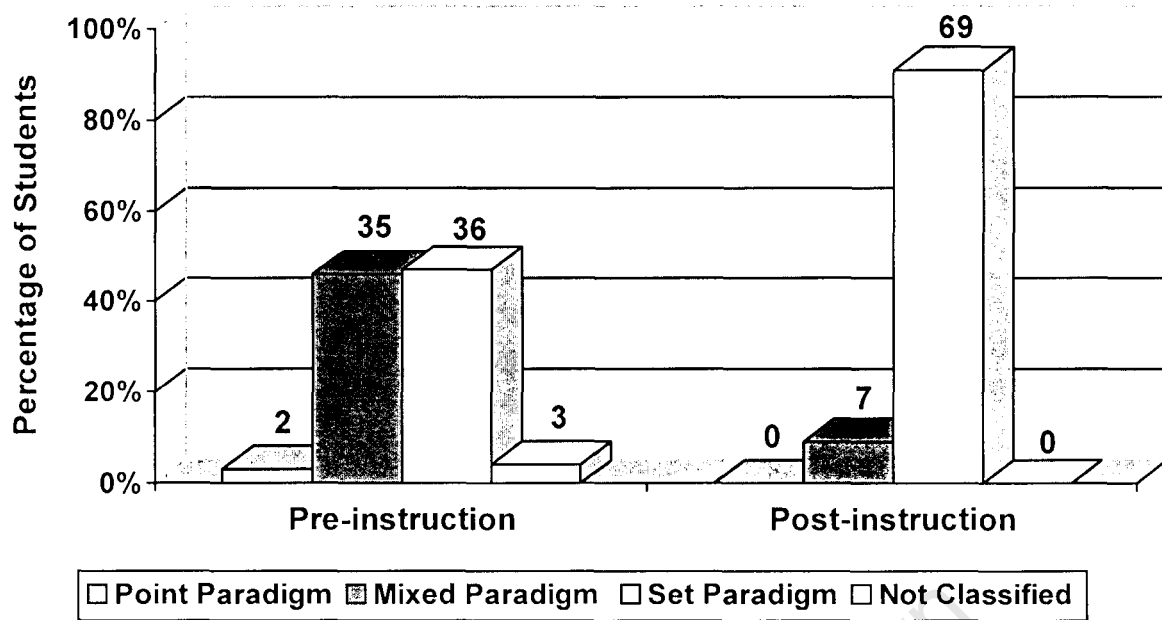


Figure 5.2: Comparison of paradigm distributions for probes relating to the average of a data set before and after instruction.

After the course, 91% of respondents consistently subscribed to set reasoning and 9% used mixed paradigms. The number of point-thinkers was reduced to zero.

5.2 Understanding of measurement in data comparison

The SMDS and DMSS probes concentrate on the comparison of two data sets, with the same average (SMDS), and with different averages (DMSS). Students were classified into either the point paradigm, set paradigm, mixed paradigm, or not classified category based on their responses to both probes.

Two set-based responses (SS) relegated students to the set paradigm category; point-thinkers were similarly identified (PP) and put in the point paradigm category; a combination of point and set responses (PS) saw students assigned to the mixed paradigm category; and the remaining response combinations (PU, SU and UU) were categorised as 'not classified'.

Table 5.3 shows the number of students in each paradigm category before instruction, and indicates the paradigms adopted by the same students after instruction.

Table 5.3: *Pre- and post-instruction paradigm distributions among students when comparing data sets ($n = 76$).*

		Post - instruction (SMDS, DMSS)				Total
		Consistent Point Paradigm	Mixed Paradigm	Consistent Set Paradigm	Not Classified	
Pre - instruction (SMDS, DMSS)	Consistent Point Paradigm	6 (8%)	19 (25%)	26 (34%)	1 (1%)	52 (68%)
	Mixed Paradigm	0 (0%)	5 (7%)	8 (11%)	0 (0%)	13 (17%)
	Consistent Set Paradigm	0 (0%)	1 (1%)	2 (3%)	0 (0%)	3 (4%)
	Not Classified	2 (3%)	3 (4%)	1 (1%)	2 (3%)	8 (11%)
Total		8 (11%)	28 (37%)	37 (49%)	3 (4%)	76 (100%)

Prior to instruction, 17% of respondents based their responses to each of the data comparison probes on a different paradigm. A review of the responses revealed that when given two data sets with the same average (SMDS), students looked to the spread of the data sets for an indication of agreement between the sets and the quality of the measurement, while they considered only the average value when presented with two data sets with different averages (DMSS).

Given that students at the beginning of the year had no previous instruction in data comparison, it is not unexpected that they did not suggest comparison of data sets based on the overlap of intervals. More than two thirds (68%) of students based their comparisons of the data sets on individual readings in the sets, or on the average alone. These students were classified as point paradigm users.

Upon completion of the course, 5 of the 13 pre-instruction mixed paradigm users retained their views, while 8 shifted to the set paradigm category. The number of students employing mixed paradigms increased from 13 to 28. In contrast to the pre-course mixed paradigm users, the larger part (59%) of the post-course group provided point-based responses to the SMDS probe and set-based responses to the DMSS probe. The implication in these cases was that the relative spreads and overlaps of the data sets were relevant only when the averages were different.

Of the 52 pre-course point reasoners, 12% persisted in their point-based reasoning and 37% used mixed paradigms after the course, while 50% demonstrated a switch to the set paradigm.

Figure 5.3 offers an overall impression of the change in students' paradigm distributions before and after instruction.

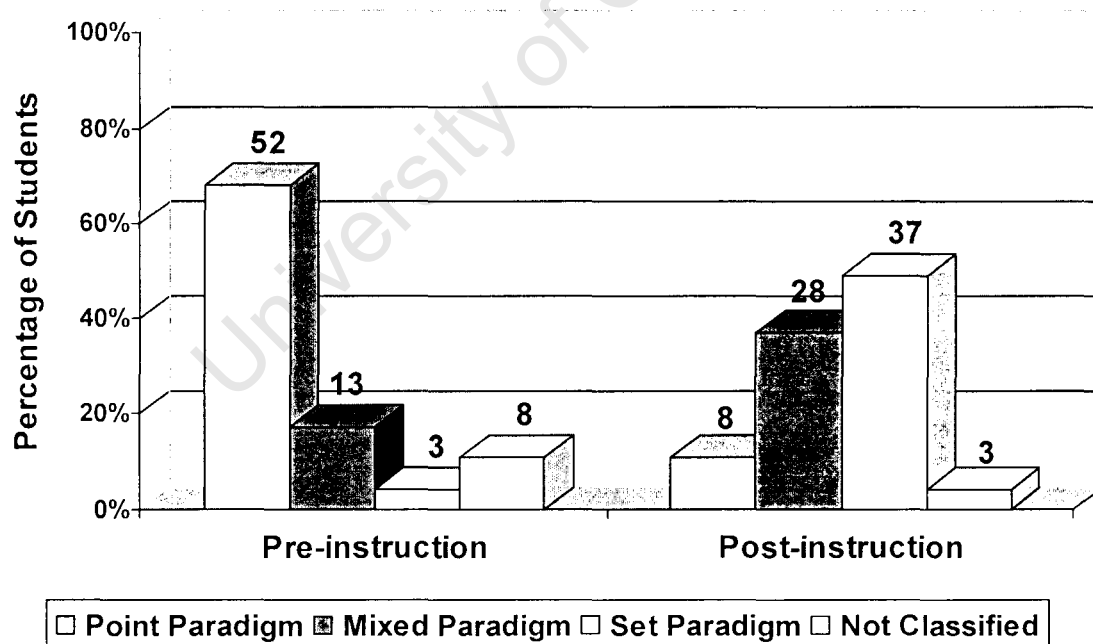


Figure 5.3: Comparison of paradigm distributions for probes relating to comparing data sets before and after instruction.

While 49% of students were consistently able to reason according to the set paradigm after instruction, 51% were still unable to appreciate the application and interpretation of the data analysis tools presented in the course.

5.3 Understanding measurement uncertainty

The NU1 probe investigated students' understanding of measurement prior to the course, and the PX1 probe undertook the same task after the course. The relative shifts in the paradigms on which students' based their responses are presented in Table 5.4.

Table 5.4: *Pre- and post-instruction paradigm distributions among students when considering uncertainty in measurement ($n = 76$).*

		Post - instruction (PX1)			Total
		Point Paradigm	Set Paradigm	Not Classified	
Pre - instruction (NU1)	Point Paradigm	3 (4%)	23 (30%)	0 (0%)	26 (34%)
	Set Paradigm	2 (3%)	45 (59%)	0 (0%)	47 (62%)
	Not Classified	0 (0%)	2 (3%)	1 (1%)	3 (4%)
Total		5 (7%)	70 (92%)	1 (1%)	76 (100%)

The set paradigm category housed the majority of students (62%) before the course, with 34% falling into the point paradigm category and 4% in the unclassifiable category.

The respondents in the set paradigm category adopted set-based ideas when presented with the NU1 probe - given the choice between the possibility of finding the true value through careful work and practice, and the impossibility of finding the true value, students chose the latter, citing various uncontrollable factors that contribute uncertainties to the measurand.

At the end of the course, in response to the PX1 probe, again questioning whether uncertainty can be reduced to zero, 23 of the 26 students in the point paradigm category moved to the set paradigm category.

The overall paradigm shifts are shown in Figure 5.4. The 92% of students identified as set-thinkers, post-instruction, attests to the success of the course in imparting to the students, an understanding of the irreducibility of uncertainties associated with measurement.

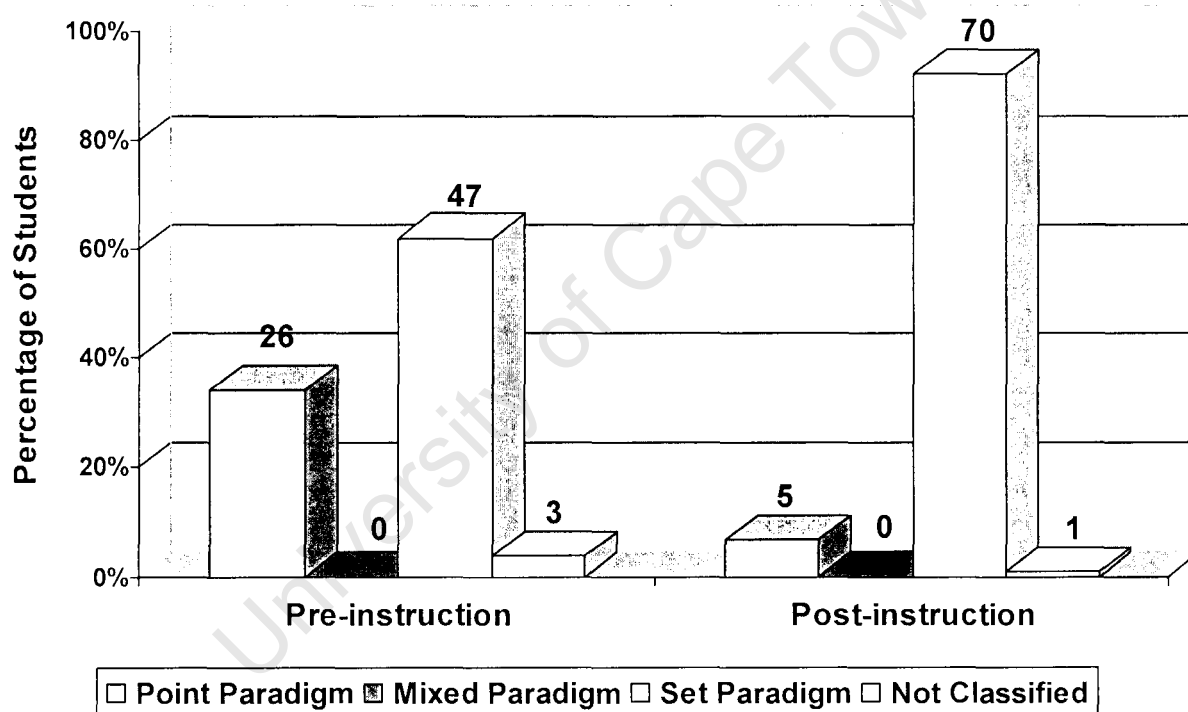


Figure 5.4: Comparison of paradigm distributions for probes related to uncertainty in measurement before and after instruction.

5.4 Cross-probe analysis for all probes

Students were classified into either the point, set or mixed paradigm category based on their responses to all nine probes common to both questionnaires. The probes covered all areas of measurement except data collection since the RD and RDA probes were included in the pre-intervention questionnaire only.

Consistent set-based responses (seven of nine) relegated students to the set paradigm category, point-thinkers were similarly put in the point paradigm category, and a combination of point and set responses saw students assigned to the mixed paradigm category.

The results of the pre-post comparison of students' paradigm choices are presented in Table 5.5.

Table 5.5: *Pre- and post-instruction paradigm distributions among students across all probes (n = 76).*

		Post - instruction			Total
		Consistent Point Paradigm	Mixed Paradigm	Consistent Set Paradigm	
Pre - instruction	Consistent Point Paradigm	0 (0%)	2 (3%)	2 (3%)	4 (5%)
	Mixed Paradigm	0 (0%)	14 (18%)	54 (71%)	68 (89%)
	Consistent Set Paradigm	0 (0%)	1 (1%)	3 (4%)	4 (5%)
Total		0 (0%)	17 (22%)	59 (78%)	76 (100%)

Upon entry to the course, only 5% of the student cohort was identified with the set paradigm of thought, 5% fell into the consistent point paradigm category and the majority (89%) was associated with mixed paradigm use. Of the latter group, 79% (54 of 68) had changed the basis of their reasoning from mixed paradigms to the consistent use of the set

paradigm by the end of the course. With 4% of students offering consistently set-based responses before and after participation in the course, and 3% progressing from point paradigm use to set paradigm use, the total number of students demonstrating a view of measurement and uncertainty grounded in the set paradigm, upon completion of the laboratory course, was 59 (78% of students). The change in paradigm distributions before and after the course is highlighted in Figure 5.5.

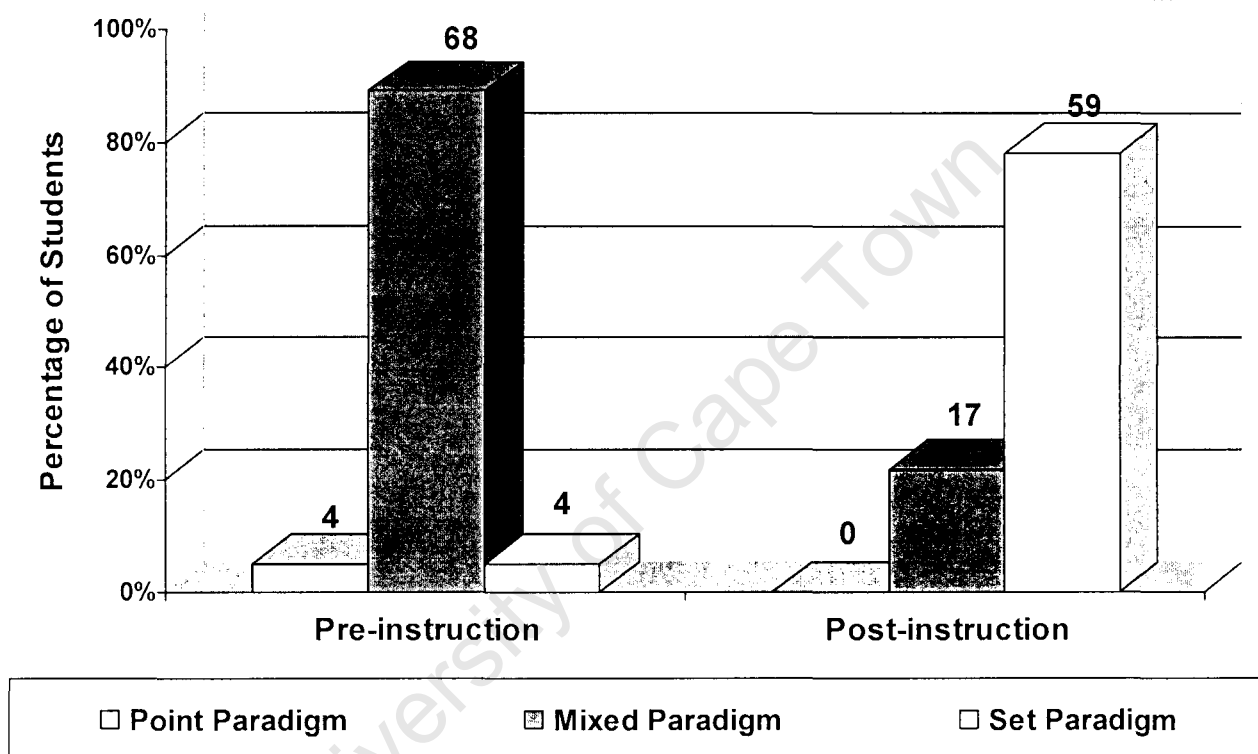


Figure 5.5: Comparison of paradigm distributions across all probes before and after instruction.

Chapter 6: Discussion

6.1 Critical reflections on the research methods

The data collection methods used in this study are presented in detail in Chapter 3. Careful attention was paid to the processes of developing the data collection instrument, to collecting the data, to ensuring the reliability of the data collected, and to analysing the data. The validity of the questionnaires was enhanced by the inclusion of existing probes which were used and validated in previous studies; the conditions under which the questionnaires were administered (as detailed in section 3.5) contributed to validating the data collection process; and the inclusion of multiple probes addressing the same area of measurement increased the reliability of the responses obtained. The validity of the analysis process was increased by the involvement of three independent researchers in the developing of the new coding schemes, and in the coding of the responses.

While these measures were carefully employed to obtain valid and reliable results, some shortcomings were observed in the process of obtaining the results. These occurrences were however, minimal, and had little impact on the analysis process and results.

Despite a description of the experimental context with an accompanying illustration on the cover of the questionnaire, and a demonstration using a scaled-up version of the apparatus, a few responses, to the QD probes, betrayed a limited grasp of the ‘rolling ball’ experimental context. Students’ were uncertain about what distance was represented by ‘ d ’. This led to an unusually high proportion of unclassifiable responses for this probe. Also, some responses to the NU1 probe indicated a misread of the stem. The underlying constructs needed to be inferred from the justifications provided. A clearer illustration or a more detailed demonstration might have averted such misunderstandings.

Although asked to provide full, independent answers to each probe, a small number of students referred to previous responses in their explanations. The phrases “as I said

before” and “same as the previous one” were observed. Again, greater emphasis on this instruction is a possible remedy.

It is important to note that the majority of students involved in this study do not speak English as a first language. Although this was taken into account in designing the questionnaires and formulating the instructions, it could be a contributor to some misinterpretations of the instructions and probe questions.

The sample cohort for the study participated voluntarily. For the pre-post comparisons the actual sample consisted of around half of the total population of all GEPS students. There is no information on the relative composition of the ‘missing half’ in terms of their previous laboratory experience, aptitude, ability or language competency. As Volkwyn (2005) has shown, previous laboratory experience, in particular, may have an influence on pre-instruction views of measurement and uncertainty. Thus the composition of the sample prompts caution in generalizing the results to the total GEPS student cohort.

The analysis method adopted was based on grouping responses according to their underlying constructs – the analysis categories emerged from the data. Simultaneously, the point and set paradigm framework was used to categorize these constructs. In principle, this is a high-risk strategy as constructs may not neatly fit in the framework.

However, various studies involving undergraduate students in both physics, (Lubben *et al.*, 2001; Buffler *et al.*, 2001; Buffler *et al.*, 2003; Lippman, 2003; Volkwyn, 2005), and chemistry, (Davidowitz *et al.*, 2001; Rollnick *et al.*, 2001; Rollnick *et al.*, 2002), have attested to the viability of the point and set paradigm framework as a basis for the classification of students’ measurement decisions at different stages of experimentation. This study lends further confirmation to the same.

6.2 Answers to the research questions

The above-mentioned caveats notwithstanding, this study has yielded some important findings.

6.2.1 Students' perceptions of measurement and uncertainty

The first of the two research questions posed in this study seeks to highlight students' understanding of measurement and uncertainty in various experimental contexts. The results show that the ideas expressed can largely be categorized according to either the point paradigm or the set paradigm depending on which particular construct serves as the basis for their interpretation of the measurement context.

The responses to the data collection probes, RD and RDA, which were administered prior to the start of the laboratory course only, support previous findings (Buffler *et al.*, 2003; Rollnick *et al.*, 2002) that entering university students generally perceive measurement in science as the pursuit of 'exact' answers or 'true' values. As with the subjects of Séré *et al.*'s (1993) investigation, the repetition of measurements is viewed as a means of confirming a previous reading, or of finding a recurring reading among students involved in this study. This reasoning is indicative of the point paradigm, which is also the underlying precept motivating, on average, 38% of students to suggest that a single reading can be chosen to represent a collection of data points.

The majority of responses to the data processing probes dealing with repeated readings are relatively easily identifiable as set paradigm-based. This trend is evident before and after the laboratory course intervention. Prior to instruction, students tend to represent sets of readings by the average value. However, Buffler *et al.*'s (2001) suggestion that this may be a result of rote-learned procedures taught at schools rather than any deeper understanding, lends some ambiguity to this paradigm classification. Post-intervention responses suggest a greater appreciation of the role of the average together with the standard deviation as a representation of a data set, and are thus rooted in the set paradigm.

Students' pre-intervention perceptions of single readings are firmly point paradigm-based. Their idea that sensitive enough apparatus will yield a single reading which is the true value of a measurand supports the findings of Evangelinos *et al.* (1998). This notion persists among 25% of the sample cohort, post-intervention. The majority of post-intervention responses, however, alludes to set paradigm-based reasoning. Students display a higher level of understanding in their assertion that single readings are best estimates of the measurand, with associated uncertainties which can be modeled by appropriate probability density functions. Not unexpectedly, some of these latter-mentioned responses seem merely rote repetitions of concepts introduced in the laboratory course.

When required to compare data sets, before participation in the laboratory course, close to three out of four respondents base their decisions about the compatibility of the sets on the averages only, and less than a quarter consider the spreads of the data sets relevant. This is in direct contrast to the results reported by Volkwyn (2005) on the study of mainstream students, which indicated that, although lacking an appreciation of the significance of the data spread to the measurement result, students were aware of the relevance of the spread as a indicator of the quality of the data set, even before instruction. This can be attributed to the educational disadvantages of the GEPS students in the study sample, who have little or no experience in laboratory exercises. Post-instruction, more than half the respondents shifted to employing the set paradigm as the foundation for their responses, which suggest not only consideration of the spreads of the data values, but also of the calculated averages and standard uncertainties, when comparing data sets. However, about one third of students, while able to carry out set-based actions like using overlapping intervals to compare data sets, were identified as point-reasoners in the cross-probe analysis.

Students' point paradigm-based expectation of a single perfect reading and one true value of a measurand, as noted by Fairbrother and Hackling (1997), is re-iterated in responses to the probes specifically aimed at eliciting students' understanding of uncertainty. The prevailing belief, before instruction, is that the uncertainty in measurement can be eliminated through careful attention to enhancing experimental skills and advancing the quality of experimental apparatus. After the course, an overwhelming shift in perceptions

from the point paradigm to the set paradigm is evidenced by students subscribing to the view that uncertainties associated with measurement cannot be reduced to zero.

6.2.2 Effectiveness of the new laboratory course

The second research question deals with the effectiveness of the new laboratory course in shifting students' perceptions of measurement and uncertainty from point to set paradigm-based. This shift is investigated in the areas of data processing with respect to single readings and multiple readings, data comparison, perceptions of uncertainty and across all measurement areas.

Students' confusion when dealing with single readings was clearly evident before their participation in the laboratory course, with almost half of respondents simultaneously employing the point and set paradigms and 5% offering either no response or a response with an unidentifiable underlying precept. Only about one tenth of students displayed consistent set-based reasoning, while more than a third consistently associated with the point paradigm. After the course, the set-based responses increased from a little more than one tenth to just less than half, while point paradigm use decreased to a mere 5%. While these paradigm shifts are notable, the large number of respondents classified in the mixed paradigm category post-intervention, are indicative of some persisting confusion.

A relatively high percentage of students applied set paradigm-based reasoning to their measurement decisions when dealing with collections of data points, even before any intervention. Upon completion of the course this percentage increased to 90.8%, and the number of consistently point-based responses was reduced to zero.

Post intervention, slightly less than half of the students consistently based their actions and reasoning on the set paradigm across all data comparison probes. Prior to instruction, more than two thirds of the sample cohort were identified as consistent point paradigm-reasoners. This number was reduced to just 8 (of 76) after instruction.

Students' paradigm use when contemplating uncertainty in measurement, before instruction, saw a third of respondents consistently associated with the point paradigm,

and close to half the sample cohort identified as set-reasoners. After the laboratory course, almost all respondents had shifted to the set paradigm.

The shift in students' paradigm use exhibited across all measurement probes was notable. The percentage of students associated with consistent set paradigm-based reasoning increased from 5% prior to the course to 78% after the course, with the percentage of respondents originally subscribing to both paradigms, decreasing from 89% to 22%, and the number of students in the point paradigm reduced to zero.

6.3 Findings related to previous studies

6.3.1 Comparison of the effectiveness of the old and new laboratory courses

The original GEPS laboratory course was evaluated by Buffler *et al.* (2001). The study focused on the responses of 70 GEPS students to written probes administered at the beginning of the year, and after a 12-week run of the original GEPS laboratory course (Allie and Buffler, 1998). The evaluation is summarized by Volkwyn (2005) and in Campbell *et al.*'s (2005) monograph. Data extracted from these sources is used to compare the proficiency of the old laboratory course in effecting paradigm shifts, from the point paradigm to the set paradigm, in students' understanding of measurement and uncertainty, to that of the new laboratory course.

Data processing:

Figure 6.1 presents a comparison of students' abilities to process repeated readings after completing the old (frequentist) GEPS course and the new (probabilistic) GEPS course, respectively.

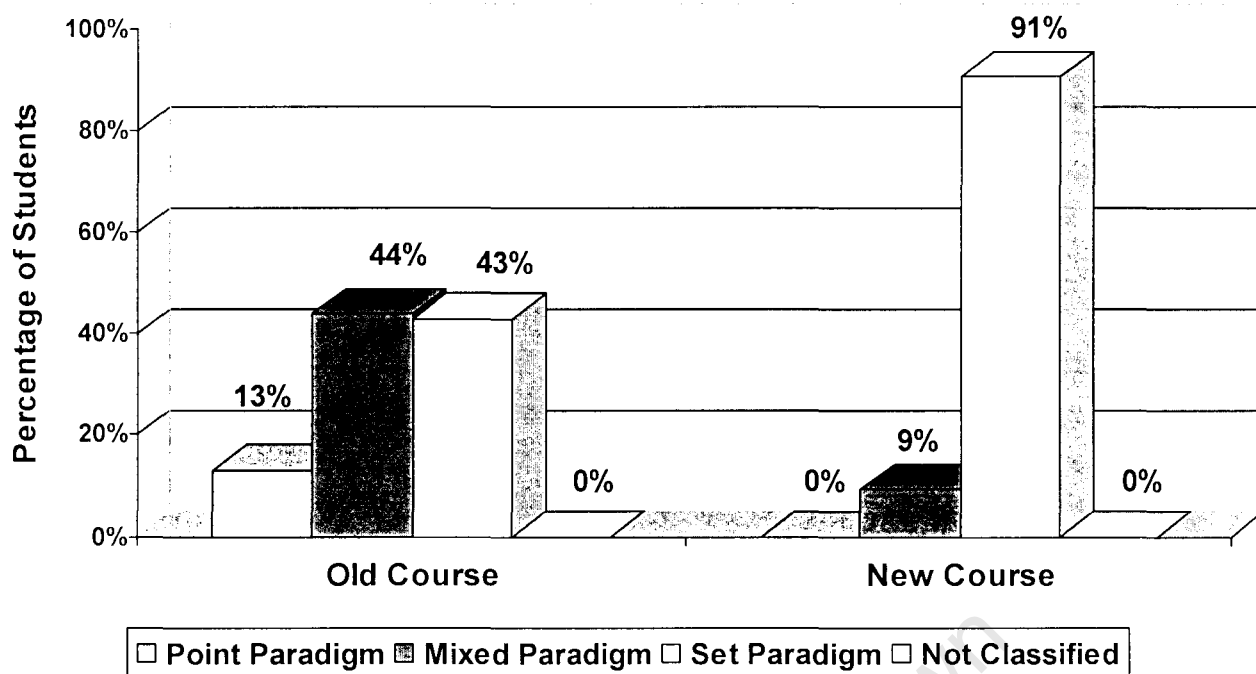


Figure 6.1: Students' paradigm use when processing data after participation in the old and new laboratory courses.

As presented in Figure 6.1, participation in the old laboratory course saw 43% of students surveyed, adopt consistent set paradigm-based actions when processing data sets. A similar percentage (44%) displayed inconsistent paradigm use, subscribing to both point and set reasoning, while only 13% indicated reasoning rooted in the point paradigm. While some improvement in students' understanding is evident for the old course, the shift pales in significance when compared to the paradigm changes effected by the new laboratory course in students' handling of multiple readings. After participation in the course, all but one in ten of the students surveyed had consistently adopted set paradigm reasoning.

Data comparison:

The data comparison probes have proved fundamental to discerning whether or not students' actions are consistent with their reasoning. The old laboratory course proved ineffectual in improving students' understanding of the relevance of the spread of data, and the standard deviation of the mean, to the comparison of data sets. An overwhelming majority of respondents (98%) employed both the point and set paradigms interchangeably

before the course, and more than two thirds (70%) of respondents retained mixed paradigm use after the course. When presented with the means and standard deviations of data sets, students judged the compatibility of the sets by considering the overlap of the intervals; however, when presented with average only, data sets were compared by their averages only, or by individual readings. This suggests that the original GEPS laboratory course was able to impart the formalistic rules of data comparison (overlapping intervals), without improving students' fundamental understanding of the statistical nature of measurement.

The new laboratory course appears to be more successful in this area. The post-instruction responses reveal that although more than a third of students demonstrated inconsistent paradigm use, close to half of the sample cohort had shifted to consistent set paradigm reasoning after the course.

Students' use of paradigms when comparing data sets after participation in the laboratory courses is illustrated in Figure 6.2.

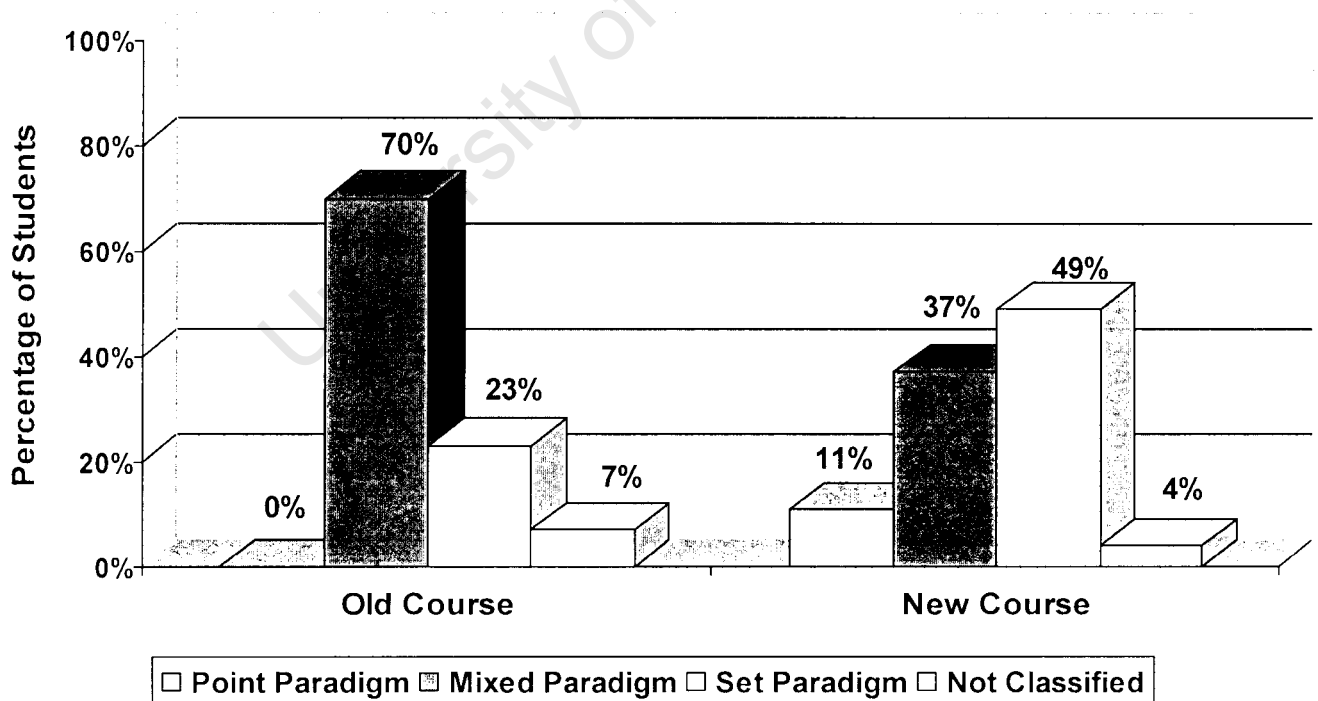


Figure 6.2: Students' paradigm use when comparing data after participation in the old and new laboratory courses.

The original GEPS laboratory course saw similar numbers of students consistently using the set paradigm (30) and displaying inconsistent use of paradigms (29) after participation, further indication that the course was not particularly successful in shifting students' perception of scientific measurement from the point paradigm view to the set paradigm one. In contrast, after participation in the new laboratory course, 59 of the 76 students in the sample cohort were classified as consistent set paradigm users, with the remaining 6 classed in the mixed paradigm category.

Figure 6.3 shows students' post-instruction paradigm use across all measurement areas for both the old and new laboratory course.

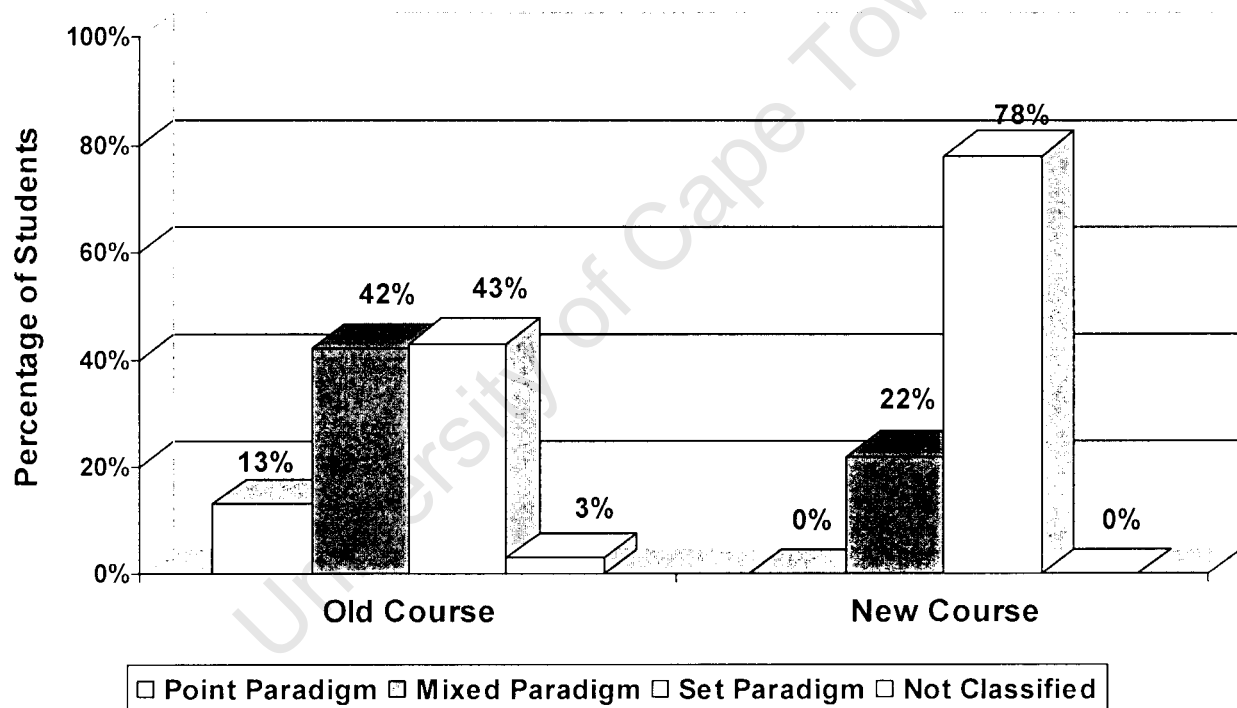


Figure 6.3: Students' paradigm use across all measurement areas after participation in the old and new laboratory courses.

An overview of the paradigms adopted by students after instruction in the old and new laboratory courses is presented in Table 6.1.

Table 6.1: *Comparison of paradigms used by students after participation in the original and new GEPS laboratory courses.*

		Post - instruction Paradigms			
		Consistent Point Paradigm	Mixed Paradigm	Consistent Set Paradigm	Not Classified
Processing data sets	Old	13%	44%	43%	0%
	New	0%	9%	91%	0%
Comparing data sets	Old	0%	70%	23%	7%
	New	11%	37%	49%	4%
All measurement areas	Old	13%	42%	43%	3%
	New	0.0%	22%	78%	0 %

6.4 Why the new course is better

The new introductory physics laboratory course evaluated here, follows the ISO-advocated probabilistic approach to teaching measurement and uncertainty. This approach offers a consistent method for dealing with single and multiple measurements, and a consistent language for communicating measurement results.

The course is administered as a combination of ‘hands-on’ laboratory activities and written exercises. It is completely interactive; allowing students to engage directly with experimental scenarios and apparatus through the laboratory activities, with the concepts of measurement and uncertainty through the written exercises, and with each other by working in small groups.

During laboratory activities students are given experimental tasks presented in plausible 'real-life' scenarios. They are required to plan and execute experimental investigations to address the given tasks. When completing written exercises, students' views are solicited and questioned, and they are forced to explore and question their own understanding. In both situations, students are encouraged to discuss the issues arising with other group members, and roving tutors are on hand to offer assistance where required.

The exercises and activities take into account students' existing beliefs about the nature of measurement in everyday and scientific contexts, and challenge their point perceptions of scientific measurement. Students are then guided towards a set paradigm-based view of measurement and uncertainty, and are explicitly introduced to the concepts underlying the various aspects of measurement, and to the statistical analysis of measurement.

Studies carried out by Buffler *et al.* (2001), Buffler *et al.* (2003) and Volkwyn (2005) suggest that the primary aim of introductory physics laboratory courses should be to change students' use of paradigms when making measurement decisions, from the point paradigm to the set paradigm.

The combination of a probabilistic approach to teaching using interactive materials, drawing on, and challenging, students' existing views, and explicitly addressing issues of how to deal with measurement and uncertainty, render the new course highly successful in effecting an overall shift in paradigms adopted in various stages of measurement, from the point paradigm to the set paradigm.

The paradigm shifts observed across the probes dealing with processing data sets, and across the probes concerning the understanding of uncertainty, suggest that the new laboratory course is proficient in communicating not only the fundamental procedures of data set analysis, but also the concepts informing the analysis, as well as an idea of the nature of measurement uncertainty.

Majority paradigm shifts, from the point paradigm to the set paradigm, are also observed across the probes dealing with processing single measurements, and those focusing on

comparing data sets. However, the change in adopted paradigms in these areas of measurement is significantly less pronounced than that in other areas.

6.5 Implications for teaching

The findings of this study support those of Leach (1999), Ryder and Clarke (2001) and Osbourne (2003), each of which claims that the explicit teaching of various aspects of experimental procedures and analysis results in improvement in students' understanding in the particular area covered. Osbourne's (1996) proposal that 'hands-on' laboratory activities should involve students in all aspects of experimental tasks from formulating and executing experimental investigations, to analysing and interpreting experimental data, is also supported by the results of this work.

The results further suggest that a combination of factors need to be taken into account to successfully impart to students, an internalised understanding of measurement and uncertainty. These factors include students' epistemological beliefs (Hammer, 1994, Elby, 2000), their views about measurement and uncertainty (Séré *et al.*, 1998), and their perceptions of the purposes of experimental tasks (Millar *et al.*, 1994), all of which affect students' success in learning and understanding physics.

The implication is that an interactive teaching approach, based on the probabilistic framework, which takes into account students' existing beliefs and explicitly addresses the various aspects of measurement and uncertainty, appears to be the route to adopt to attain the teaching and learning objectives (described in section 2.1) of the introductory physics laboratory.

Also suggested by the findings of this study, is the need for particular focus on the areas of single measurements, as noted by Evangelinos *et al.* (1998), and the comparison of data sets (Masnick and Morris, 2002), in laboratory work.

6.6 Further work

The focus of this work has been a new probabilistic laboratory course which is the product of a ten-year collaborative research project involving investigations into students' understanding of measurement and uncertainty, and the evaluations of other introductory physics laboratory courses. The evaluation completed in this study reveals the overall pedagogic success of this course. The findings of the evaluation also point out measurement areas, viz. comparing data sets and processing single measurements, in which the course has limited success in achieving its teaching objectives. These results prompt further questions.

- Which aspects of the course are most effective in facilitating the observed shifts, from the point paradigm to the set paradigm, in students' perceptions of measurement and uncertainty?
- Why is the course less successful in particular areas of measurement; and how can these shortcomings be remedied?
- Are there any parts of the course that are ineffectual in significantly imparting knowledge to students, or that act as barriers to the learning process? If so, how can any part of these failures be attributed to the choice of language and vocabulary employed in the workbook?

Implementing further investigation of these questions will allow for the modification and amendment of the new laboratory course in order to enhance its strengths and reduce its deficiencies.

References

- Abbott, D. S. (2003) *Assessing Student Understanding of Measurement and Uncertainty*
Unpublished PhD thesis: North Carolina State University.
- Allie, S. and Buffler, A. (1998) A course in tools and procedures for Physics I. *American Journal of Physics*, 66 (7), 613-624.
- Allie, S., Buffler, A., Campbell, B., Lubben, F., Evangelinos, D., Psillos, D. and Valassiades, O. (2003) Teaching measurement in the introductory physics laboratory. *The Physics Teacher*, 41 (7), 394-401.
- Allie, S., Buffler, A., Kaunda, L. and Ingles, M. (1997) Writing-intensive physics laboratory reports: tasks and assessment. *The Physics Teacher*, 35, 399-405.
- Allie, S., Buffler, A., Kaunda, L., Campbell, B. and Lubben, F. (1998) First year physics students' perceptions of the quality of experimental measurements. *International Journal of Science Education*, 20 (4), 447-459.
- Black, P. (1993) The purposes of science education. In R. Hull (Ed.), *ASE Secondary Science Teachers' Handbook*. London: Simon & Schuster.
- Buffler, A., Allie, S., Lubben, F. and Campbell, B. (2001) The development of first year physics students' ideas about measurement in terms of point and set paradigms. *International Journal of Science Education*, 23 (11), 1137-1156.
- Buffler, A., Allie, S., Lubben, F. and Campbell, B. (2002) *Introduction to measurement in the physics laboratory. A probabilistic approach*. Laboratory Manual, Department of Physics, University of Cape Town.
(Can be downloaded from <http://www.phy.uct.ac.za/people/buffler/labmanual.html>)

- Campbell, B., Lubben, F., Buffler, A. and Allie, S. (2005) *Teaching Scientific Measurement at University: understanding students' ideas and laboratory curriculum reform*. (Monograph endorsed by African Journal of Research in Mathematics, Science and Technology Education). SAARMSTE.
- Chi, M., Feltovitch, P. and Glaser, R. (1981) Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Coelho, S. and Séré, M-G. (1998) Pupils' reasoning and practice during hands-on activities in the measurement phase. *Research in Science and Technological Education*, 16 (1), 79-96.
- Cohen, L. and Manion, L. (1980) *Research Methods in Education*. 241-262 Newbury Park: Sage.
- d'Agostini, G. (1999) *Bayesian Reasoning in High Energy Physics – Principles and Applications*. (CERN Yellow Report 99-3). Geneva: CERN.
- Davidowitz, B., Lubben, F. and Rollnick, M. (2001) Undergraduate science and engineering students' understanding of the reliability of chemical data. *Journal of Chemical Education*, 78 (2), 247-252.
- Deardoff, D. L. (2001) *Introductory Physics Students' Treatment of Measurement Uncertainty* Unpublished PhD thesis: North Carolina State University.
- Elby, A. (2001) Helping physics students learn how to learn. *American Journal of Physics*, 69 (7), S54-S64.
- Etkina, E., Van Heuvelen, A., Brookes, D. and Mills, D. (2002) Role of experiments in physics instruction - a process approach. *The Physics Teacher*, 40, 351-355.

- Evangelinos, D., Psillos, D. and Valassiades, O. (1998) Students' introduction to measurement concepts: A metrological approach. In European Commission Report on Project PL 95-2005 *Labwork in Science Education*. pp. 561-587.
- Evangelinos, D., Psillos, E. and Valassiades, O. (2002) An investigation of teaching and learning about measurement data and their treatment in the introductory physics laboratory. In D. Psillos and H. Niederrerr (Eds.): *Teaching and learning in the science laboratory*. Dordrecht: Kluwer Academic Publishers. pp. 179-190.
- Fairbrother, R. and Hackling, M. (1997) Is this the right answer? *International Journal for Science Education*, 19 (8), 887-894.
- Garrett, J., Horn, A. and Tomlinson, J. (2000) Misconceptions about error. *University Chemistry Education*, 4 (2), 54-57.
- Gee, B. and Clackson, S. G. (1992) The origin of practical work in the English school science curriculum. *School Science Review*, 73(265), 79 - 83.
- Germann, P., Aram, R. and Burke, G. (1996) Identifying patterns and relationships among the responses of seventh-grade students to the science process skill of designing experiments. *Journal of Research in Science Teaching*, 33 (1), 79-99.
- Germann, P. and Aram, R. (1996) Student performances on the science processes of recording data, analysing data, drawing conclusions and providing evidence. *Journal of Research in Science Teaching*, 33 (7), 773-798.
- Gillham, B. (2000) *Developing a Questionnaire*. 2-15 Continuum.
- Giordano, J. G. (1999) On reporting uncertainties of the straight line regression parameters, *European Journal of Physics*, 20 (5), 343-349.
- Gott, R. and Duggan, S. (1996) Practical work: its role in the understanding of evidence in science. *International Journal of Science Education*, 18 (7), 791-805.

- Hammer, D. (1994) Epistemological beliefs in introductory physics. *Cognition and Instruction*, 12, 151-183.
- Hammer, D. and Elby, A. (2003) Tapping epistemological resources for learning science. *Journal of the Learning Sciences*, 12 (1), 53-90.
- Hodson, D. (1998) Taking practical work beyond the laboratory. *International Journal of Science Education*, 20 (6), 629-632.
- International Organization for Standardization (1993) International Vocabulary of Basic and General Terms in Metrology (VIM). Geneva: ISO.
- International Organization for Standardization (1995) Guide to the expression of uncertainty in measurement (GUM). Geneva: ISO.
- Kirschner, P. and Huisman, W. (1998) Dry laboratories in science education: computer-based practical work. *International Journal of Science Education*, 20 (6), 665-682.
- Kuhn, D., Amsel, E. and O'Loughlin, M. (1998) *The development of scientific thinking skills..* London: Academic Press.
- Larkin, J. and Reif, F. (1979) Understanding and teaching problem solving in physics, *European Journal of Science Education*, 1, 191-203.
- Laws, P. (1996) Millikan Lecture 1996: Promoting active learning based on physics education research in introductory physics courses. *American Journal of Physics*, 65 (1), 14-21.
- Leach, J. (1999) Students' understanding of the co-ordination of theory and evidence in science. *International Journal of Science Education*, 21 (8), 789-806.

- Leach, J., Millar, R., Ryder, J. and Séré, M-G. (2000) Epistemological understanding in science learning: the consistency of representations across contexts. *Learning and Instruction*, 10, 497-527.
- Lippmann, R. (2003) *Students' understanding of measurement and uncertainty in the physics laboratory: social construction, underlying concepts, and quantitative analysis*. Unpublished PhD thesis: University of Maryland.
- Lubben, F. and Millar, R. (1996) Children's ideas about the reliability of experimental data. *International Journal of Science Education*, 18, 955-968.
- Lubben, F., Buffler, A., Allie, S. and Campbell, B. (2001) Point and set reasoning in practical science measurement by entrant university freshmen. *Science Education*, 85, 311-327.
- Masnick, A. and Morris, B. (2002) Reasoning from data: the effect of sample size and variability on children's and adults' conclusions. In W. Gray and C. Schunn (Eds.): *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum. pp. 643-648.
- McGinn, M. and Roth, W-M. (1999) Preparing students for competent scientific practice: implications of recent research in science and technology studies. *Educational Researcher*, 28 (3), 14-24.
- Meester, M. and Maskill, R. (1995) First year chemistry practicals at universities in England and Wales - aims and the scientific level of the experiments. *International Journal of Science Education*, 17 (5), 575-588.
- Millar, R. (1996) Student investigations in science: a knowledge-based approach. *Didaskalia*, 9, 9-30.

- Millar, R., Le Marechal, J-F. and Tiberghien, A. (1999) 'Mapping' the domain: varieties of practical work. In D. Psillos and H. Niederrerr (Eds.): *Teaching and learning in the science laboratory*. Dordrecht: Kluwer Academic Publishers. pp. 33-59.
- Millar, R., Lubben, F., Gott, R. and Duggan, S. (1994) Investigating in the school science laboratory: Conceptual and procedural knowledge and their influence on performance. *Research Papers in Education*, 9, 207-248.
- Montes, L. and Rockley, M. (2002) Teacher perceptions in the selection of experiments. *Journal of Chemical Education*, 79 (2), 244-247.
- Osborne, J. (1996) Untying the Gordian knot: diminishing the role of practical work. *Physics Education*, 31 (5), 271-278.
- Pfundt, H. and Duit, R. (1994) *Bibliography: Students' Alternative Frameworks and Science Education*. Kiel: IPN.
- Reif, F. and Larkin, J. H. (1991) Cognition in scientific and everyday domains: comparison and learning implications. *Journal of Research in Science Teaching*, 28 (9), 733-760.
- Rollnick, M., Dlamini, B., Lotz, S. and Lubben, F. (2001) Views of South African chemistry students in university bridging programmes on the reliability of experimental data. *Research in Science Education*, 31 (4), 553-573.
- Rollnick, M., Lubben, F., Lotz, S. and Dlamini, B. (2002) What do under prepared students learn about measurement from introductory laboratory work? *Research in Science Education*, 32 (1), 1-18.
- Roth, W. and Roychoudhury, A. (1993) The development of science process skills in authentic contexts. *Journal of Research in Science Teaching*, 30, 127-152.

- Roth, W-F., McRobbie, C., Lucas, K. and Boutonne, S. (1997) Local production of order in traditional science labs. *Learning and Instruction*, 8 (2), 107-136.
- Ryder, J. and Clarke, A. (2001) Teaching and Learning about “Sources of Error” on University Physics Courses *A Research Study supported by the Learning and Teaching Support Network - Physical Sciences*, University of Leeds
- Ryder, J. and Leach, J. (2000) Interpreting experimental data: the views of upper secondary school and university science students. *International Journal of Science Education*, 22 (10), 1069-1084.
- Séré, M-G., Fernandez-Gonzalez, F., Gallegos, J., Gonzalez-Garcia, F., De Manuel, E., Perales, J. and Leach, J. (2001) Images of science linked to labwork: a survey of secondary school and university students, *Research in Science Education*, 31, 499-523.
- Séré, M-G., Journeaux, R. and Larcher, C. (1993) Learning the statistical analysis of measurement error. *International Journal of Science Education*, 15 (4), 427-438.
- Song, J. and Black, P. (1992) The effect of concept requirements and task contexts on pupils’ performance in control variables. *International Journal of Science Education*, 14 (1), 83-93.
- Strauss, A. and Corbin, J. (1990) *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park: Sage.
- Taylor, B. and Kuyatt, C. (1994) *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*. (NIST Technical Note 1297, 1994). Also available at <http://physics.nist.gov/Pubs/guidelines/contents.html> .
- Thomson, V. (1997) Precision and the terminology of measurement. *The Physics Teacher*, 35 (1), 15-17.

- Tiberghien, A., Veillard, L., Le Marechal, J-F., Buty, C. and Millar, R. (2001) An analysis of labwork tasks used in science teaching at upper secondary school and university levels in several European countries. *Science Education*, 85 (5), 483-508.
- Tomlinson, J., Dyson, P. and Garratt, J. (2001) Student misconceptions of the language of error. *University Chemistry Education*, 5 (1), 1-8.
- Volkwyn, T., Allie, S., Buffler, A., Lubben, F. and Campbell, B. (2004) First year physics students' understanding of measurement in the context of laboratory practicals. In A. Buffler and R. Laugksch (Eds.): *Proceedings of the 12th Annual Conference of the Southern African Association for Research in Mathematics, Science and Technology Education*. Cape Town, pp. 1011-1017.
- Volkwyn, T. S. (2005) First year Students' Understanding of Measurement in Physics Laboratory Work Unpublished MSc thesis: University of Cape Town.
- Vygotsky, L. (1978) *Mind in society*. Oxford: Blackwell.
- White, R. T. (1996) The link between the laboratory and learning. *International Journal of Science Education*, 18 (7), 761-774.

Appendix I

The complete pre-intervention measurement questionnaire

University of Cape Town

Surname:

First name:

K

Jan 2003

University of Cape Town
Department of Physics

Physics Measurement Questionnaire

Instructions:

Write your name in the box above.

Inside this envelope there are pages numbered up to page 13.

Read the text below and answer the questions on each sheet.

If you need more space for your answers, then use the backs of the sheets.

It should take you about 5 minutes to answer each question.

Answer the questions in order and do not skip any sheet.

When you have completed a question, put the sheet inside this envelope and do not take it out again, even if you want to change your answer.

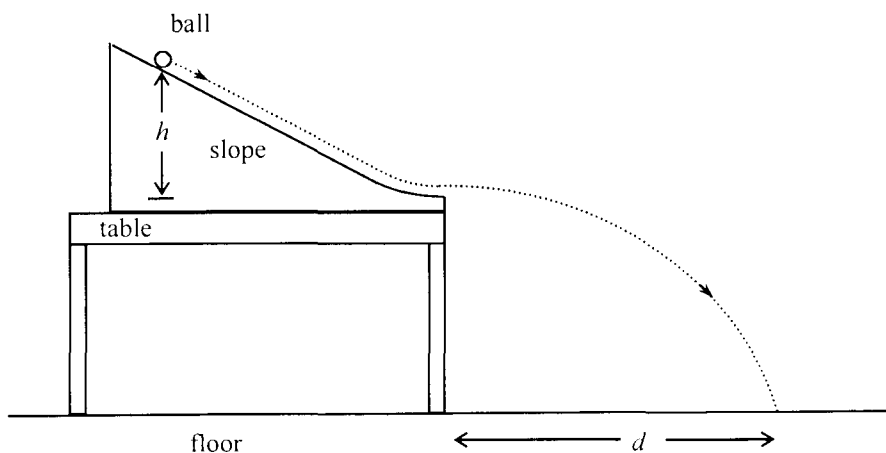
Note: It is possible that some answers may be similar or exactly the same as others. Please write all answers out in full, even if you feel that you are repeating yourself.

Context:

An experiment is being performed by students in the Physics Laboratory.

A wooden slope is clamped near the edge of a table. A ball is released from a height h above the table as shown in the diagram. The ball leaves the slope horizontally and lands on the floor a distance d from the edge of the table. Special paper is placed on the floor on which the ball makes a small mark when it lands.

The students have been asked to investigate how the distance d on the floor changes when the height h is varied. A metre stick is used to measure d and h .



Q 1. (QD1/K)

The students work in groups on the experiment. Their first task is to determine d when $h = 90$ mm. One group lets the ball roll down the slope from a height $h = 90$ mm and use a metre rule to measure the distance d . What they see is shown below.



Spot made by the ball
on the paper

I think that the distance d is exactly 436 mm.



A

I think that the distance d is approximately 436 mm.



B

I think that the distance d is between 435 mm and 437 mm.



C

I think that the distance d is between 435.5 and 436.5 mm.



D

I don't agree with any of you.



E

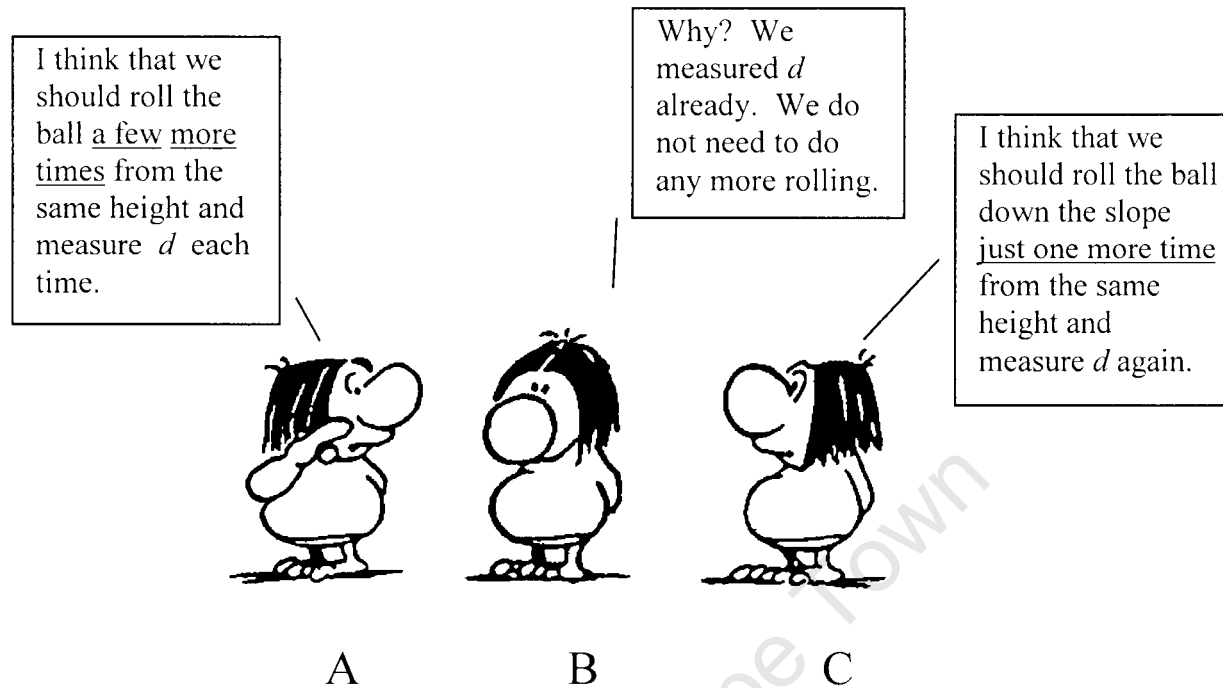
With whom do you most closely agree? (Circle ONE):

A	B	C	D	E
---	---	---	---	---

Explain your choice.

Q 2. (RD/K)

The following discussion now takes place between the students.



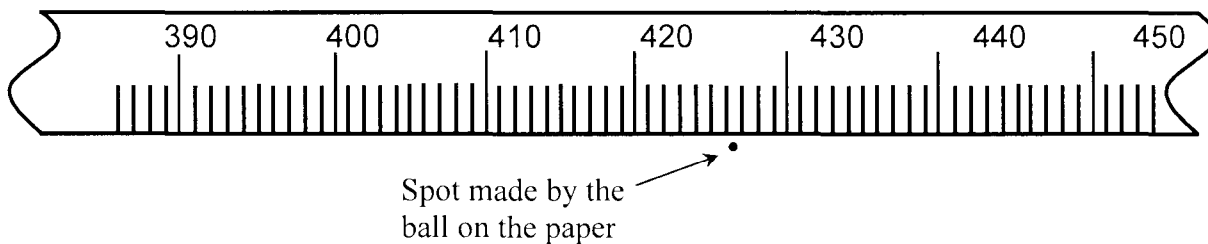
With whom do you most closely agree? (Circle ONE):

A	B	C
---	---	---

Explain your choice.

Q 3. (QD2/K)

The group of students decide to allow the ball to roll again from height $h = 90$ mm. The students use the same metre rule to measure the distance d , and what they see is shown below.



I think that the distance d is exactly 426.5 mm.



A

I think that the distance d is approximately 426.5 mm.



B

I think that the distance d is between 426 and 427 mm.



C

I don't agree with any of you.



D

With whom do you most closely agree? (Circle ONE):

A	B	C	D
---	---	---	---

Explain your choice.



Q 4. (RDA/K)

After two rolls from the same height of $h = 90$ mm, the students have the following readings:

First release: $h = 90$ mm $d = 436$ mm

Second release: $h = 90$ mm $d = 426$ mm

The following discussion then takes place between the students.

We know enough.
We don't need to roll the ball again.

We need to roll the ball just one more time.

Three rolls will not be enough. We should roll the ball several more times and measure d each time.



A

B

C

With whom do you most closely agree? (Circle ONE):

A	B	C
---	---	---

Explain your choice.

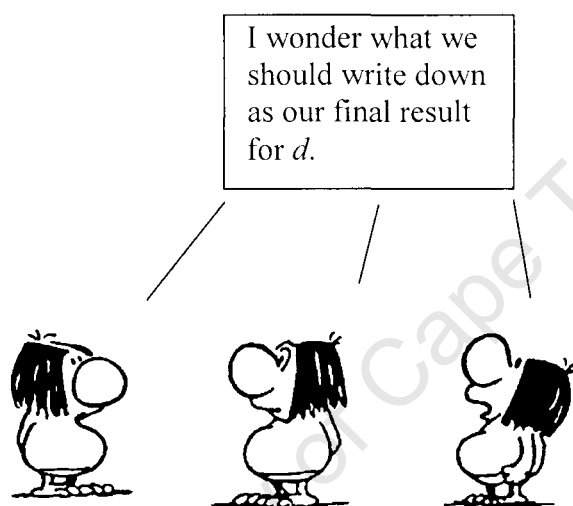


Q 5. (UR1/K)

The students continue to allow the ball to roll down the slope from the same height $h = 90$ mm. Their readings after five rolls are:

<u>Roll</u>	<u>d (mm)</u>
1	436
2	426
3	434
4	430
5	434

The students then discuss what to write down for d as their final result.



Write down what you think the students should write down as their final result for d .

Explain your answer.

Q 6. (UA1/K)


The students decide to calculate the average of their readings for d , which is 432 mm.

<u>Roll</u>	<u>d (mm)</u>
1	436
2	426
3	434
4	430
5	<u>434</u>

Average: **432**


They then discuss what the average for the distance d tells them.

I think that the distance d is exactly 432 mm.




A

I think that the distance d is approximately 432 mm.




B

I think that the distance d is somewhere between 431.5 mm and 432.5 mm.




C

I think that the distance d is somewhere between 426 mm and 436 mm.



D

I don't agree with any of you.



E

With whom do you most closely agree? (Circle ONE):

A	B	C	D	E
---	---	---	---	---

Explain your choice.

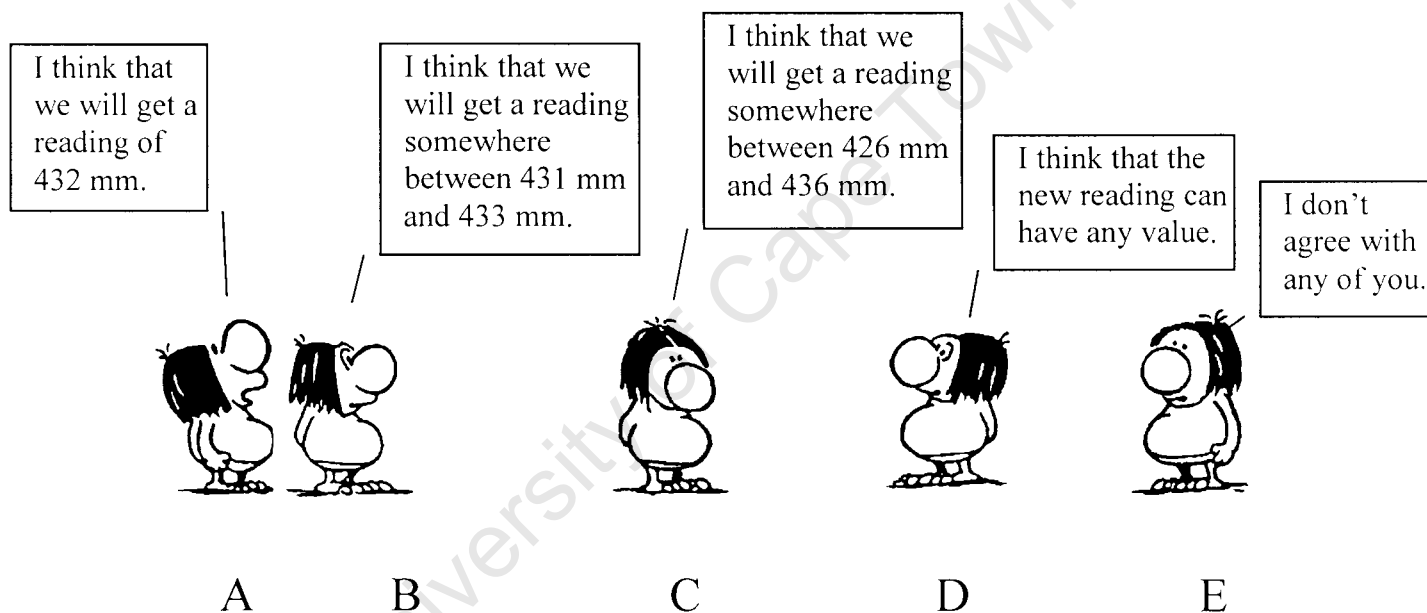
Q 7. (UA2/K)

The students have 5 readings for d obtained from allowing the ball to roll from the same height $h = 90$ mm:

<u>Roll</u>	<u>d (mm)</u>
1	436
2	426
3	434
4	430
5	<u>434</u>

Average: **432**

The students now discuss what reading they will get for d if they roll the ball again (for the sixth time) from $h = 90$ mm.



With whom do you most closely agree? (Circle ONE):

A	B	C	D	E
---	---	---	---	---

Explain your choice.

Q 8. (SMDS/K)

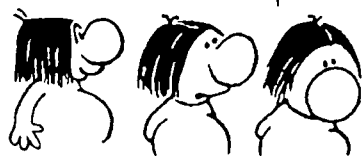
Two groups of students compare their measurement of d obtained by letting the ball roll from $h = 90$ mm. Their readings for five rolls are shown below, together with their averages.

Roll	Group A d (mm)	Group B d (mm)
1	444	441
2	432	460
3	424	410
4	440	424
5	435	440
Average:	435	435

Our result for d is better. All our readings are between 424 mm and 444 mm. Your readings are spread between 410 mm and 460 mm.

Our result for d is just as good as yours. Our average is the same as yours. We both got 435 mm for d .

I think that the result of group B is better than the result of group A.



A



B



C

With which group do you most closely agree? (Circle ONE):

A

B

C

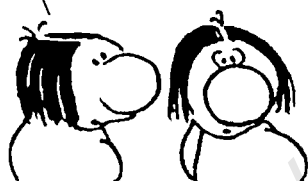
Explain your choice. Do not use the word "results" in your explanation.

Q 9. (DMSS/K)

Two other groups of students compare their measurement of d obtained from allowing the ball to roll from $h = 90$ mm. Their readings for five rolls are shown below, together with their averages.

Roll	Group A d (mm)	Group B d (mm)
1	440	432
2	438	444
3	433	426
4	422	433
5	432	440
Average:	433	435

Our result for d
agrees with yours.



A

No, our results
do not agree.



B

With which group do you most closely agree? (Circle ONE):

A

B

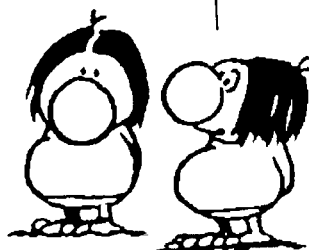
Explain your choice. Do not use the word "results" in your explanation.



Q 10. (NU1/K)

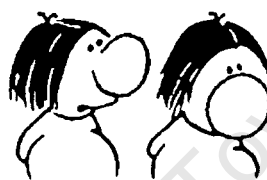
When they are finished, the two groups discuss how they can improve their rolling ball experiment next time.

If we practice enough and work very carefully, all our readings will be the same. Then we will know the true value of d .



A

No, even if your readings are all the same, you will still not know the true value of d .



B

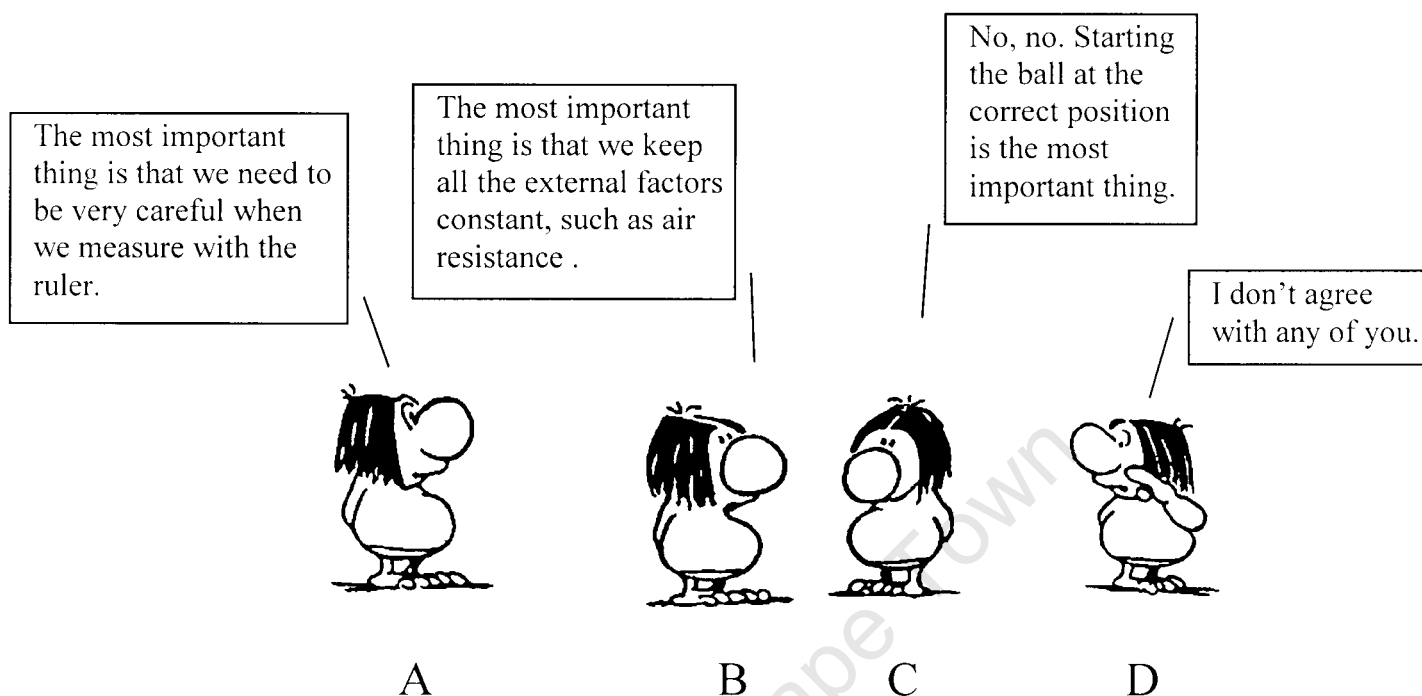
With which group do you most closely agree? (Circle ONE):

A	B
---	---

Explain your choice.

Q 11. (NU2/K)

The students continue to discuss how to improve their experiment next time.



With whom do you most closely agree? (Circle ONE):

A	B	C	D
---	---	---	---

Explain your choice.

Q 11B. (AE1/K)

The following discussion then takes place between some of the students.

It is ok to have approximate measurements for everyday use but for physics, the measurements have to be exact.



A

It all depends on whether you want to find a mathematical formula or not.



B

No, all measurements are always approximate.



C

No, all measurements are always exact.



D

I don't agree with any of you.



E

With whom do you most closely agree? (Circle ONE):

A

B

C

D

E

Explain your choice.



Q 12. (QD3/K)


The lecturer now comes around with a special electronic meter which has a digital display and uses it to measure the distance d for one of the rolls from $h = 90$ mm.

Here is what the electronic meter shows:




The following discussion takes place between the students.

I think that the distance d is exactly 423.7 mm.




A

I think that the distance d is approximately 423.7 mm.




B

I think that the distance d is between 423 mm and 424 mm.




C

I think that the distance d is between 423.65 and 423.75 mm.



D

I don't agree with any of you.

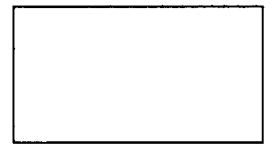


E

With whom do you most closely agree? (Circle ONE):

A	B	C	D	E
---	---	---	---	---

Explain your choice.



Q 13.

Comments.

Are there any answers to the previous question sheets that you want to change?

Please do not remove any sheets from the envelope.

What was the question about and how do you want to change your answer?



Any other comments?



In this laboratory questionnaire, I thought that the cartoon figures were (tick one):	male	female	mixed gender
---	------	--------	--------------

Appendix II

The coding schemes for the pre-intervention probes

For each coding scheme, the table headings are:

Number : The number of student responses allocated to each category.

Code : The alphanumeric code for each category

P/S : The allocation of each category to either the point or set paradigm, if appropriate.

Category : A short description of each category.

University of Cape Town

Q 1. (QD1/K):

Number Code P/S Category

0	N00	-	No response
0	U00	-	Not able to code response
	A		exactly 436 mm, because ...
0	A00	-	(no reason given)
0	A01	-	(not able to code reason given)
0	A10	P	of the size of the spot, doesn't extend beyond mark.
0	A11	P	the ball has travelled exactly 436 mm.
10	A12	P	the spot is exactly on the 436 mm mark.
6	A13	P	the distance is easily read/seen/observed from scale on ruler.
4	A20	P	measurement is exact.
0	A41	P	exact conditions for repeating will result in the exact same distance.
	B		approximately 436mm, because ...
0	B00	-	(no reason given)
4	B01	-	(not able to code reason given)
3	B10	P	of the size of the spot.
4	B11	P	of the size/shape/movement of ball.
2	B12	P	the spot is not exactly on a mark.
1	B13	P	closer markings/smaller intervals are needed, scale is too coarse.
2	B14	S	human judgement is required to estimate reading, error of parallax.
1	B15	P	a better measuring instrument is needed.
1	B16	P	of the mass of the ball.
2	B17	U	the distance the ball travels is the same as/similar to the height from which it is released.
			the total distance the ball travels includes the distance along the slope, the fall
0	B18	U	from the table, and the distance covered between landing on the ground and stopping.
1	B21	S	measurement/the reading is not exact/perfect, so use 'approximate'.
1	B22	S	Physics, (and/or Maths), requires exact numbers, but measurements are not exact, so use 'approximate'.
0	B23	P	the spot/dot is 'more or less' at 436 mm.
0	B30	U	need to repeat measurements.
1	B31	P	repeating will give a recurring value.
2	B32	S	repeating will allow calculation of average.
1	B33	S	repeating will confirm range of measurements.
4	B40	S	the distance/position of spot is influenced by external factors.
1	B41	S	the distance/position of spot is influenced by experimental procedure/measurement process.

C between 435 mm and 437 mm, because ...

0 C00 - (no reason given)

1 C01 - (not able to code reason given)

3 C10 P of the size of the spot.

3 C11 P of the size/shape/movement of ball.

1 C12 P the spot is not exactly on a mark.

0 C16 P of the mass of the ball.

0 C17 U the distance the ball travels is the same as/similar to the height from which it is released.
the total distance the ball travels includes the distance along the slope, the fall

1 C18 U from the table, and the distance covered between landing on the ground and stopping.

1 C41 S of experimental conditions.

D between 435.5 mm and 436.5 mm, because ...

0 D00 - (no reason given)

1 D01 - (not able to code reason given)

1 D10 P of the size of the spot.

2 D11 P of the size/shape/movement of ball.

0 D12 P the spot is not exactly on a mark.

1 D13 P closer markings/smaller intervals are needed, scale is too coarse.

0 D16 P of the mass of the ball.

0 D17 U the distance the ball travels is the same as/similar to the height from which it is released.
the total distance the ball travels includes the distance along the slope, the fall

0 D18 U from the table, and the distance covered between landing on the ground and stopping.

1 D20 S measurement is always uncertain, can't be sure.

E I don't agree with any of you, because ...

0 E00 - (no reason given)

2 E01 - (not able to code reason given)

0 E10 P of the size of the spot.

1 E16 P of the mass of the ball.

1 E17 U the distance the ball travels is the same as/similar to the height from which it is released.
the total distance the ball travels includes the distance along the slope, the fall

2 E18 U from the table, and the distance covered between landing on the ground and stopping.

0 E20 S measurement is always uncertain, can't be sure.

1 E21 S there is no exact answer.

 E30 U repeats are necessary.

1 E40 S of external factors.

1 E41 S of experimental conditions.

Q 2. (RD/K):

Number	Code	P/S	Category
0	N00	-	No response
0	U00	-	Not able to code response
A			
I think we should roll the ball <u>a few more times</u> from the same height and measure d each time because ...			
0	A00	-	(no reason given)
2	A01	-	(not able to code reason given)
6	A10	P	practice will produce a more accurate or better measurement.
0	A11	P	practice will reduce the systematic error in the measurement
0	A12	P	you have to repeat until the readings are close together
21	A20	S	you need more readings to get an average/mean
1	A21	S	you need to get a more accurate/reliable average/mean
0	A22	S	you need to get an average and a spread/uncertainty
0	A23	S	you need to get an average and a better/narrower spread/uncertainty
0	A24	S	you need to get an average in order to get closer to the true value
35	A30	P	a few more rolls may get you the same (i.e. correct) answer
5	A40	P	you need to get a variety of results
1	A60	P	you have to do it several times (no reason provided)
0	A62	P	you must always take three measurements
8	A64	P	the answer gets more accurate; closer to the true value
0	A72	S	you need to determine the spread/uncertainty
0	A73	S	you need to determine a better/narrower spread/uncertainty
0	A74	P	you need to determine the uncertainty to get closer to the true value
B			
Why? We've got the result already. We do not need to do any more rolling, because			
0	B00	P	(no reason given)
0	B01	P	(not able to code reason given)
2	B30	P	repeating will give the same result
0	B40	P	repeating will give different results which is confusing
0	B50	P	repeating is a waste of time or resources
C			
I think we should roll the ball down the slope <u>just one more time</u> from the same height, because ...			
0	C00	-	(no reason given)
0	C01	-	(not able to code reason given)
2	C10	P	practice will make the second measurement more accurate
0	C11	P	practice will reduce the systematic error in the measurement
0	C20	S	you can calculate the average from two measurements
0	C21	S	you can get a more accurate/reliable average
15	C30	P	you need to see if you get the same (i.e. correct) result
2	C40	P	you need to get a variety of results
0	C50	P	many repeats are a waste of time or resources
0	C51	P	many repeats are desirable, but time consuming

Q 3. (QD2/K):

Number	Code	P/S	Category
0	N00	-	No response
0	U00	-	Not able to code response
	A		exactly 426.5 mm, because ...
0	A00	-	(no reason given)
2	A01	-	(not able to code reason given)
3	A12	P	the spot is exactly in the middle.
1	A41	P	exact conditions for repeating will result in the exact same distance.
	B		approximately 426.5 mm, because ...
0	B00	-	(no reason given)
1	B01	-	(not able to code reason given)
4	B10	P	of the size of the spot.
0	B11	P	of the size/shape/movement of ball.
2	B12	P	the spot is not exactly on a mark.
4	B13	P	closer markings/smaller intervals are needed, scale is too coarse.
0	B14	S	human judgement is required to estimate reading, error of parallax.
0	B15	P	a better measuring instrument is needed.
1	B21	S	measurement/the reading is not exact/perfect, so use 'approximate'.
2	B22	S	in Science/Physics/ Maths measurements are not exact, so use 'approximate'.
5	B23	P	the spot/reading is 'more or less' at 436 mm, so use 'approximate'.
0	B31	P	must repeat to find accurate answer.
1	B32	S	repeating will allow calculation of average.
0	B33	S	repeating will confirm range of measurements.
2	B40	S	the distance/position of spot is influenced by external factors.
4	B41	S	the distance/position of spot is influenced by experimental procedure/measurement process.
4	B50	S	it is the number that best represents the interval.
0	B51	S	the measurement is approximate with respect to the true value and experimental error.
	C		between 426 mm and 427 mm, because ...
0	C00	-	(no reason given)
1	C01	-	(not able to code reason given)
3	C10	P	of the size of the spot.
3	C11	P	of the size/shape/movement of ball.
13	C12	P	the spot is not exactly on a mark.
5	C13	P	closer markings/smaller intervals are needed, scale is too coarse, need better calibration.
0	C40	S	external factors cause variations when repeating.
4	C50	S	the measurement is in the interval.

	D	I don't agree with any of you, because ...
0	D00	- (no reason given)
4	D01	- (not able to code reason given)
1	D11	P of the movement/size/shape of the ball.
0	D30	U need to repeat measurements.
1	D32	S must repeat to calculate average.
0	D35	S repeat to get best approximation
1	D40	S of external factors
3	D41	S of experimental conditions/mistakes.
1	D60	U must round to the nearest 10.

University of Cape Town

Q 4. (RDA/K):

Number	Code	P/S	Category
0	N00	-	No response
0	U00	-	Not able to code response
	A		We know enough. We don't need to repeat the measurement again, because ...
0	A00	P	(no reason given)
0	A01	P	(not able to code reason given)
1	A40	P	repeating will give a different result again, which is confusing
1	A50	P	it saves doing it again; repeats are a waste of time / resources
	B		We need to release the ball <u>just one more time</u>, because
0	B00	P	(no reason given)
0	B01	P	(not able to code reason given)
0	B10	P	practice will make the third measurement even more accurate
2	B11	P	practice will reduce the systematic error in the measurement
8	B20	S	you need more measurements to get an average / mean
0	B21	S	you need to get a more accurate average / mean
0	B22	S	you need to get an average and a spread / uncertainty
0	B23	S	you need to get an average and a more accurate / narrower uncertainty
0	B24	S	you need to get the average in order to get closer to the true value
4	B30	P	the 3rd measurement may give the same (i.e. correct) answer
3	B31	P	you need to find a pattern in the readings.
0	B40	P	3 measurements are enough; too many different answers are confusing
1	B50	P	many repeats are a waste of time or resources
0	B51	P	many repeats are desirable, but time consuming
1	B60	P	you have to do it three times (no reason provided)
2	B64	P	the answer gets more accurate; closer to the true value
0	B72	S	you need to determine the spread / uncertainty
0	B73	S	you need to determine a more accurate / narrower spread / uncertainty
0	B74	P	you need to determine the uncertainty to get closer to the true value

C		Three releases will not be enough. We should release the ball <u>several more times</u> , because	
0	C00	-	(no reason given)
2	C01	-	(not able to code reason given)
3	C10	P	the more practice, the more accurate your measurement gets
3	C11	P	practice will reduce the systematic error in the measurement
0	C12	P	you have to repeat until the measurements are close together
22	C20	S	you need more measurements to get an average / mean
3	C21	S	you need to get a more accurate average / mean
0	C22	S	you need to get the average/mean and the spread/uncertainty
0	C23	S	you need to get the average and a more accurate spread/uncertainty
0	C24	S	you need to get an average in order to get closer to the true value
12	C30	P	a few more times may get you the same (i.e. correct) answer
7	C31	P	you need to find a pattern in the readings.
8	C40	P	you need to get a large variety of results
2	C60	P	you have to do it more than three times (no reason provided)
15	C64	P	the answer gets more accurate; closer to the true value
0	C72	S	you need to determine the spread / uncertainty
0	C73	S	you need to determine a more accurate / narrower spread / uncertainty
0	C74	P	you need to determine the uncertainty to get closer to the true value
0	C80	-	you need many repeated measurements for plotting a graph

Q 5. (UR/K):

Number	Code	P/S	Category
0	N00	-	No response
6	U00	-	Not able to code response
33	10	S	average of the readings is final result for d
1	12	P	xxx mm is final result, close to average
1	20	S	average of the readings, excluding the lowest reading, is final result for d
0	30	P	median reading is final result of d
7	40	S	interval is final result for d
2	50	P	final result is first, last, highest, lowest, reading
26	60	P	final result is the recurring reading

University of Cape Town

Q 6. (UA1/K):

Number	Code	P/S	Category
0	N00		No response
0	U00		Not able to code response
	A		<i>d</i> is exactly 432 mm, because ...
0	A00		(no reason given)
0	A01		(not able to code reason given)
1	A10	P	average is the true distance
2	A30	P	average is exactly 432 mm.
2	A31	P	average is calculated using a formula, and is therefore correct.
1	A32	P	average remains the same after more repeats.
1	A41	P	if individual readings differ, then the average is the exact value.
	B		<i>d</i> is approximately 432 mm, because ...
0	B00		(no reason given)
0	B01		(not able to code reason given)
2	B10	S	average is closest to exact value/true distance.
4	B11	S	average is most likely value/best approximation of distance <i>d</i> .
7	B12	S	average is not necessarily the exact/actual/true distance.
3	B13	S	actual distance could be more or less than the average distance.
0	B16	S	average is best estimate, can never know true value.
5	B20	S	Error must be accounted for.
0	B21	S	External factors must be accounted for.
8	B30	S	average is approximate, not exact.
1	B31	S	more repeats are necessary to find accurate average/answer.
1	B33	S	average will/may change after more repeats.
4	B41	S	individual readings/measurements are not identical, so there is no exact value.
2	B42	S	readings are close to 432 mm.
2	B44	S	no reading is 432 mm.
0	B50	P	size/shape/movement of ball
0	B71	S	don't know exact interval in which <i>d</i> lies, must calculate standard uncertainty of average.
	C		<i>d</i> is between 431.5 mm and 432.5 mm, because ...
0	C00		(no reason given)
0	C01		(not able to code reason given)
0	C11	S	average is most likely value/best approximation of <i>d</i> , but not exact.
1	C13	S	actual distance is close to the average.
0	C14	S	actual distance is in interval.
1	C15	S	average is actual distance rounded off/ interval accounts for rounding off.
0	C16	S	can never be sure of the real <i>d</i> .
0	C20	S	allow/account for error/uncertainty.
2	C30	S	average is approximate/not exact/not accurate.
1	C33	S	average can/will change after more repeats.
0	C70	S	interval allows for standard uncertainty of average, (calculate to find best estimate and standard uncertainty)

	D	d is between 426 mm and 436 mm, because ...
0	D00	(no reason given)
2	D01	(not able to code reason given)
3	D11	S average in interval, so actual distance is also in interval.
1	D14	S actual distance in interval, average is best representation of interval/easy number to work with.
0	D20	S account for error/uncertainty.
1	D30	S average is approximate/not exact.
0	D31	S more repeats are necessary to find accurate average/answer.
0	D33	S average will change after more repeats.
0	D34	S average is in interval.
6	D40	S all readings are in that interval.
3	D41	S readings spread between 426 mm and 436 mm, so can't know exact distance.
2	D44	S no reading is 432 mm, so actual distance is described by the interval/range.
0	D72	S interval found by calculating standard uncertainty of average, is in this interval.
	E	I don't agree with any of you, because ...
0	E00	(no reason given)
1	E01	(not able to code reason given)
0	E13	S actual distance is close to average.
1	E14	S average distance is 432 mm.
2	E16	S average provides minimal value of actual distance.
0	E30	S average is approximate, not exact.
0	E31	S need more repeats to find accurate average/answer.
0	E32	P average is the same, whatever the readings.
1	E33	S average will change after more repeats.
1	E42	S readings are close to 432 mm.
1	E60	U average and actual d are unrelated.
0	E70	S d lies in uncertainty interval about average (best estimate), must calculate.
0	E71	S can't reach conclusion, need additional information about standard uncertainty.

Q 7. (UA2/K):

Number	Code	P/S	Category
0	N00	-	No response
0	U00	-	Not able to code response
	A		reading of 432 mm, because ...
0	A00	-	(no reason given)
1	A01	-	(not able to code reason given)
2	A30	P	average remains unchanged.
	B		reading between 431 mm and 432 mm, because ...
0	B00	-	(no reason given)
1	B01	-	(not able to code reason given)
2	B31	S	average between 431 mm and 432 mm.
1	B40	S	reading must be close to average.
	C		reading between 426 mm and 436 mm, because ...
0	C00	-	(no reason given)
2	C01	-	(not able to code reason given)
23	C10	S	all readings in range 426 mm-436 mm.
1	C11	S	readings have no set pattern.
1	C12	S	readings vary/change/are not the same.
3	C20	S	same experimental conditions, readings in same range.
6	C31	S	average is between 426 mm and 436 mm.
1	C41	S	more repeats will change average.
2	C50	S	434 mm occurs twice, reading likely to be close.
	D		reading can have any value, because ...
0	D00	-	(no reason given)
0	D01	-	(not able to code reason given)
3	D11	S	readings have no set pattern, so can't predict.
4	D12	S	readings vary/change/are not all the same, so can't predict.
6	D13	S	reading can have any value, can't be predicted.
2	D14	S	reading can have any value in specified/unspecified bigger ranger.
4	D20	S	experimental factors affect readings.
5	D21	S	external factors affect readings.
0	D40	S	reading will be close to average.
2	D41	S	reading will change average, so can't predict.
	E		I don't agree with any of you, because ...
0	E00	-	(no reason given)
1	E01	-	(not able to code reason given)
0	E12	S	readings vary/change/not the same, so can't predict.
0	E20	S	experimental conditions affect readings.
2	E21	S	external factors affect readings.
1	E30	P	average remains unchanged.
0	E40	S	readings must be close to average.

Q 8. (SMDS/K):

Number Code P/S Category

0 N00 - No response
1 U00 - Not able to code response

A A's results are better, because ...
0 A00 - (no reason given)
0 A01 - (not able to code reason given)
1 A10 S they have a smaller range/spread
0 A11 S they have a smaller range/spread because of outside factors
5 A12 S they have smaller range/spread because fewer mistakes were made
2 A13 S they have smaller range/spread , therefore a more accurate/reliable average
1 A14 S they have smaller range/spread, therefore are closer to true value
0 A15 S they have smaller range/spread because group A was more skilful
1 A20 S there is less deviance from the average
0 A21 S there is less deviance from the average because of outside factors
1 A22 S less deviance from the average because of fewer mistakes made
0 A25 S less deviance from the average because group A was more skilful
0 A40 - you usually get the results so close together
0 A50 P their average (435 mm) is also one of the measurements
7 A63 S A's results are more accurate/ consistent

B B's results are just as good as A's, because ...
0 B00 - (no reason given)
0 B01 - (not able to code reason given)
0 B10 S they got more or less the same measurements
33 B20 P they have the same average
3 B21 P they have the same average although different outside factors caused deviation
0 B22 P they have the same average although mistakes caused deviation
8 B23 P they have the same average, and the spread is not important
1 B26 P they have the same average, deviation not important as expected
1 B29 P they have the same average and same number of readings
0 B30 P they have the same average, although A got 435 mm on their last measurement.
1 B60 - there is no exact answer to an experiment like this
4 B65 P the accuracy of individual readings is not under consideration, the average is important
3 B70 P it is a natural outcome of the same experiment, the spread is not important)

C I think that the results of group B are better than the results of group A because ...
0 C00 - (no reason given)
2 C01 - (not able to code reason given)
0 C10 S B's results are closer together; they don't vary as much
0 C11 S B's average is more accurate/reliable
0 C12 S B's spread is smaller, so the average is more accurate
0 C40 - you usually get the results so close together
1 C50 P A's average (435 mm) is also one of the measurements

Q 9. (DMSS/K):

Number	Code	P/S	Category
0	N00	-	No response
1	U00	-	Not able to code response
A Our results <u>agree</u> with yours, because ...			
0	A00	-	(no reason given)
1	A01	-	(not able to code reason)
2	A10	P	the readings/measurements for both sets are more or less the same
2	A12	S	the readings/measurements for both sets have the same spread
0	A13	S	the readings/measurements have an overlapping spread
17	A20	P	the averages are more or less the same
0	A21	P	the averages are more or less the same, difference due to external factors
3	A22	P	the averages are more or less the same, difference due to experimental errors
0	A24	P	the averages are more or less the same, both close to true value
3	A26	P	the averages are more or less the same as there will always be deviation
0	A30	S	the uncertainties of the averages may overlap
2	A31	S	the averages are more or less the same with similar ranges/spreads
4	A40	P	three out of five (the majority) of readings are the same
1	A41	P	group A's first reading is the same as group B's last reading
0	A50	P	if you round off the averages, then they are identical
B No, your results <u>do not agree</u> with ours, because ...			
0	B00	-	(no reason given)
2	B01	-	(not able to code response)
1	B12	S	the spreads of both sets are different
22	B20	P	The averages are different
1	B21	P	The averages are different due to different conditions/external factors
3	B22	P	The averages are different due to experimental errors
2	B25	P	The averages are different, absolute accuracy/identical results required to agree
2	B26	P	The averages are too different even though deviation is taken into consideration
0	B30	S	the averages are too far apart for the uncertainties to overlap
4	B31	P	average is different and all individual readings are not the same
1	B32	S	The spread differs between the two
0	B40	P	both groups got some different measurements
0	B50	P	if you round off the averages, then they are very different
0	B60	P	an average is only true if the average value also appears as one of the measurements
1	B61	P	an average is only true if the average value does not appear as one of the measurements
1	B80	P	group B is more accurate than group A

Q 10. (NU1/K):

Number	Code	P/S	Category
1	N00	-	No response
0	U00	-	Not able to code response
	A		If we practice enough and work very carefully, all our readings will be the same and we will know the true value of d, because ...
0	A00	-	(no reason given)
1	A01	-	(not able to code reason)
7	A10	P	“practice makes perfect.”
10	A11	P	perfecting experimental method will result in a true/accurate/best value.
3	A12	P	if external factors are taken into account, results will be exact/accurate.
8	A13	P	exact conditions for repeating experiment will give exact results.
7	A20	P	if all the readings are the same, then that recurring value is the true value of d.
	B		No, even if all the readings are the same, we will still not know the true value of d, because ...
0	B00	-	(no reason given)
2	B01	-	(not able to code response)
7	B11	S	you cannot avoid experimental mistakes.
25	B12	S	you cannot control external factors, and therefore, cannot know true value.
3	B13	S	exact conditions for repeating experiment will give exact results, BUT conditions change.
10	B21	S	results will always vary.
6	B30	S	no true value, can only approximate.
10	B31	S	no true value, use average.

Q 11. (NU2/K):

Number	Code	P/S	Category
1	N00	-	No response
0	U00	-	Not able to code response
A The most important thing is careful measurement with the ruler, because ...			
0	A00	-	(no reason given)
0	A01	-	(not able to code reason given)
3	A10	P	accurate measurement is necessary for a successful experiment/good results.
1	A11	P	incorrect reading of scale will lead to inaccurate results/invalid experiment.
B The most important thing is keeping external factors constant, because ...			
0	B00	-	(no reason given)
3	B01	-	(not able to code reason given)
15	B10	P	results will be more accurate/correct
4	B11	P	it eliminates uncertainty in the experiment
28	B20	P	external factors are the biggest contributors to incorrect/ varying results
9	B30	P	it will make finding the true value possible
C The most important thing is the starting position of the ball, because ...			
0	C00	-	(no reason given)
0	C01	-	(not able to code reason given)
2	C10	U	results/experiment will be better/more accurate
0	C11	U	different starting points give invalid results
5	C20	U	different starting points give varying results
2	C30	P	if the starting position is correct, then we will get the true/exact/right value
1	C40	U	if the starting position is correct, then we can find a more accurate average
1	C51	U	the position of the ball is the only controllable factor
1	C52	U	the position of the ball is the most difficult factor to control
D I don't agree with any of you, because ...			
0	D00	-	(no reason given)
0	D01	-	(not able to code reason given)
4	D11	U	it is important to keep all experimental conditions constant
1	D12	U	it is most important to work accurately and carefully
14	D13	U	all mentioned factors are equally important
2	D30	S	it doesn't matter since we cannot know the true value of d, always approximate.
3	D50	S	can't control all factors, readings will vary

Q 11B. (AE1/K):

Number	Code	P/S	Category
2	N00	-	No response
1	U01	-	Not able to code response
A approximate for everyday, exact for physics, because ...			
0	A00	-	(no reason given)
3	A01	P	(not able to code reason given)
9	A11	P	in Physics/Science societal consequences of inaccuracy are greater.
5	A12	P	in Physics/Science exact values are needed for calculations.
6	A13	P	in Physics/Science exact values are needed for measurements/results.
2	A14	P	in Physics/Science exact values are needed for research.
2	A20	P	Physics/Science must be accurate.
4	A30	P	exact values are needed to get the right answer/'true value'.
B depends on whether, or not, you want to find a mathematical formula, because ...			
0	B00	-	(no reason given)
6	B01	-	(not able to code reason given)
6	B10	P	it depends on the purpose of the measurement.
5	B12	P	exact numbers are needed for formulae.
2	B60	P	if the purpose is unimportant, may settle for an average instead of exact measurement.
C all measurements are always approximate, because ...			
0	C00	S	(no reason given)
1	C01	S	(not able to code reason given)
3	C10	S	in Physics/Science cannot get exact values, only approximate ones.
1	C40	S	in experiments, one always approximates
13	C41	S	one can't be sure, so measurements are always approximate
1	C51	S	good experimental method/apparatus gives close to 'true value', but not exact
4	C52	S	of calibration of measuring instruments and/or experimental method.
8	C53	S	of external factors.
D all measurements are always exact, because ...			
1	D01	P	(not able to code reason given) <i>misread, reason does not support choice</i>
E I don't agree with any of you, because ...			
0	E00	-	(no reason given)
1	E01	-	(not able to code reason given)
1	E14	P	in Physics/Science exact values are needed for research and publication.
1	E20	P	Physics/Science/Maths requires exact measurements, (can't be approximate), because of the nature of the discipline.
1	E30	P	exact values are needed to get the right answer/'true value'
3	E40	P	measurements are not always exact.
8	E51	U	different experimental conditions give different results.

Q 12. (QD3/K):

Number Code P/S Category

- 0 N00 - No response
- 0 U00 - Not able to code response

A exactly 423.7 mm, because ...

- 0 A00 - (no reason given)
- 1 A01 - (not able to code reason given)
- 2 A11 P the ball has travelled exactly 423.7 mm.
- 2 A12 P the reading is 423.7 mm.
- 25 A15 P electronic meter gives exact/accurate answers.
- 1 A41 P exact conditions for repeating will result in the exact same distance.

B approximately 423.7 mm, because ...

- 0 B00 - (no reason given)
- 2 B01 - (not able to code reason given)
- 0 B10 P of the size of the spot.
- 0 B11 P of the size/shape/movement of ball.
- 4 B12 S the reading is approximate.
- 0 B13 P meter not sensitive enough, need better calibration.
- 1 B14 S human judgement is required to estimate reading, error of parallax.
- 1 B15 P a better measuring instrument is needed.
- 6 B21 S measurement/the reading is not exact/perfect, so use 'approximate'.
- 1 B22 S in Science/Physics/Maths measurements are not exact, so use 'approximate'.
- 2 B24 S measurement close to true value, so use approximate.
- 0 B30 U need to repeat measurements.
- 0 B31 P repeating to confirm, get recurring value.
- 1 B32 S repeating to account for scatter by calculation of average.
- 1 B34 S there are variations when repeating.
- 4 B40 S variations caused by external factors.

C between 423 mm and 424 mm, because ...

- 0 C00 - (no reason given)
- 1 C01 - (not able to code reason given)
- 0 C13 P need better calibration, more sensitive.
- 2 C15 P need better measuring device, meter inaccurate.
- 2 C20 S reading is not exact, use interval.
- 0 C33 S repeats will confirm range, readings will fall in interval.
- 0 C40 S of external factors
- 0 C41 S of experimental method/measuring process.

D between 423.65 mm and 423.75 mm, because ...

- 0 D00 - (no reason given)
- 0 D01 - (not able to code reason given)
- 0 D10 P of the size of the spot.
- 1 D11 P of the size/shape/movement of ball.
- 1 D13 P closer markings/smaller intervals are needed, scale is too coarse, need better calibration
- 1 D15 P need better measuring instrument.
- 0 D20 S measurement is always uncertain, can't be sure.
- 0 D41 S of experimental method/measuring process.

E		I don't agree with any of you, because ...	
0	E00	-	(no reason given)
2	E01	-	(not able to code reason given)
1	E15	S	of mistakes of the measuring device.
1	E20	S	measurement is always uncertain, can't be sure.
0	E30	U	repeats are necessary.
2	E31	P	repeat to confirm, find recurring value.
1	E32	S	repeat to account for scatter and calculate average.
0	E34	S	repeat to account for external factors and approximate.
7	E40	S	of external factors.

University of Cape Town

Appendix III

The complete post-intervention measurement questionnaire

University of Cape Town

Surname:

Initials:

Student number:

L

Sept 2003

University of Cape Town

Department of Physics

PHY123H

Laboratory Measurement Questionnaire

16 September 2003

Instructions:

Write your name in the box above and your student number **on each page**.

Read the text below and answer the questions on each sheet.

If you need more space then use the back of the sheet.

Answer the nine questions in order and do not skip any sheet.

After completing all the questions you can go back and revise any answer.

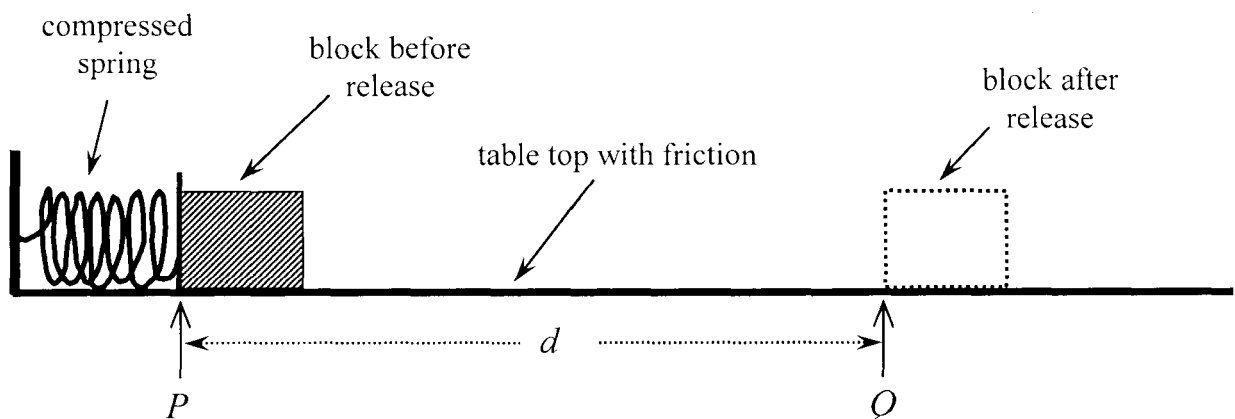
However, do **not** erase or change what you have already written. Simply write your new answer **below** the old one.

It is possible that some answers may be similar or exactly the same as others. Please write all answers out in full, even if you feel that you are repeating yourself.

If you think that an answer requires a calculation then do **not** work out the result but simply show what you would do or leave the calculation in an incomplete form.

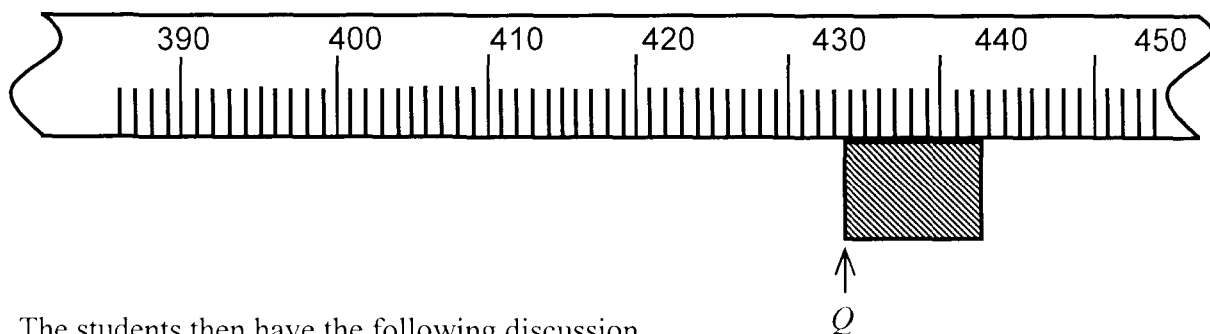
Context:

An experiment is taking place in the physics laboratory to investigate the motion of a block on a table with friction. The block is pushed against the spring so that its left edge is at position P . The block is released and travels a distance d to position Q as shown. The students, working in groups, have to determine d using a metre rule that is provided.



Q 1. (PR1/L)

One of the groups marks point P and *carefully* lines it up with the zero mark on the metre rule. They then release the block. After the block comes to rest they see that point Q lines up on the metre stick as shown.



The students then have the following discussion.

I think that the distance d the block has travelled is exactly 433.8 mm.



A

I think that the distance d the block has travelled is approximately 433.8 mm.



B

I think that the distance d the block has travelled is between 433.0 and 434.0 mm.



C

I think the distance d the block has travelled is approximately 434.0 mm.



D

I don't agree with any of you.



E

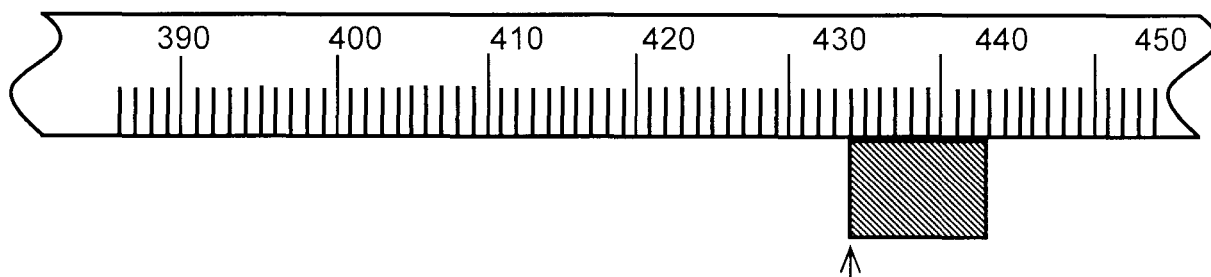
With whom do you most closely agree? (Circle ONE):

A	B	C	D	E
---	---	---	---	---

Explain your choice.

Q 2. (PR2/L)

A second group of students marks point P and *carefully* lines it up with the zero mark on the metre rule. They then release the block. After the block comes to rest they see that point Q lines up on the metre stick as shown.



The students then have the following discussion.

I think that the distance d the block has travelled is exactly 434.0 mm.



A

I think that the distance d the block has travelled is approximately 434.0 mm.



B

I think that the distance d the block has travelled is between 433.0 and 435.0 mm.



C

I think that the distance d the block has travelled is exactly 434 mm.



D

I don't agree with any of you.

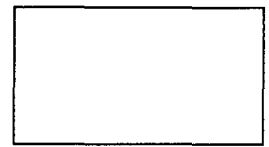


E

With whom do you most closely agree? (Circle ONE):

A	B	C	D	E
---	---	---	---	---

Explain your choice.

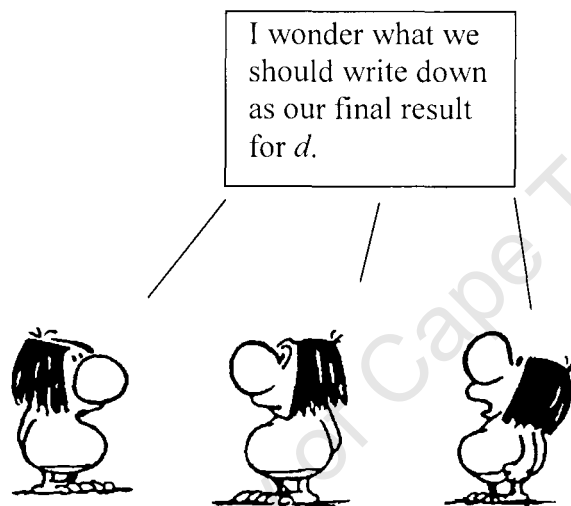


Q 3. (UR/L)

A third group of students releases the block 5 times from point P . The 5 values they obtain for d are shown below.

<u>Release</u>	<u>d (mm)</u>
1	434.5
2	426.6
3	435.8
4	432.0
5	431.1

The students then discuss what to write down as their final result for d .



Write down what you think the students should record as their final result for d and explain your answer.

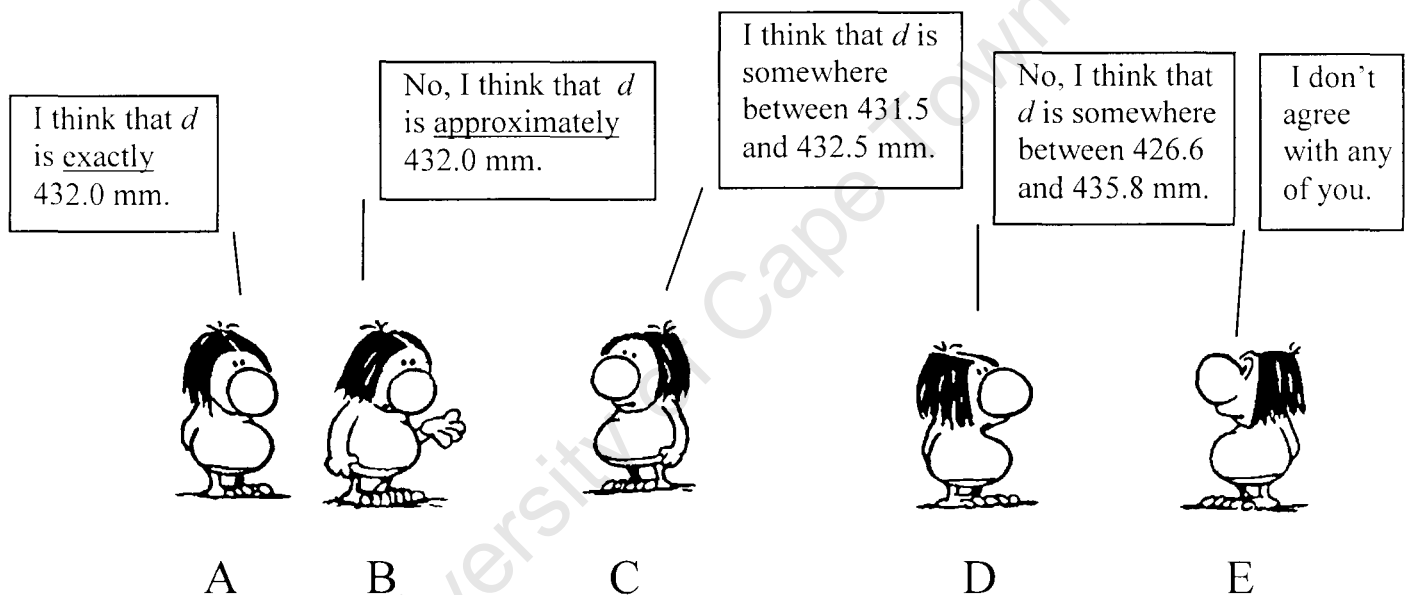
Q 4. (UA1/K)

One of the students in the last group decides to calculate the average of their readings for d , which turns out to be 432.0 mm.

Release	d (mm)
1	434.5
2	432.0
3	435.8
4	426.6
5	<u>431.1</u>

Average: **432.0**

They then discuss what they can say about d .



With whom do you most closely agree? (Circle ONE):
Explain your choice.

A	B	C	D	E
---	---	---	---	---



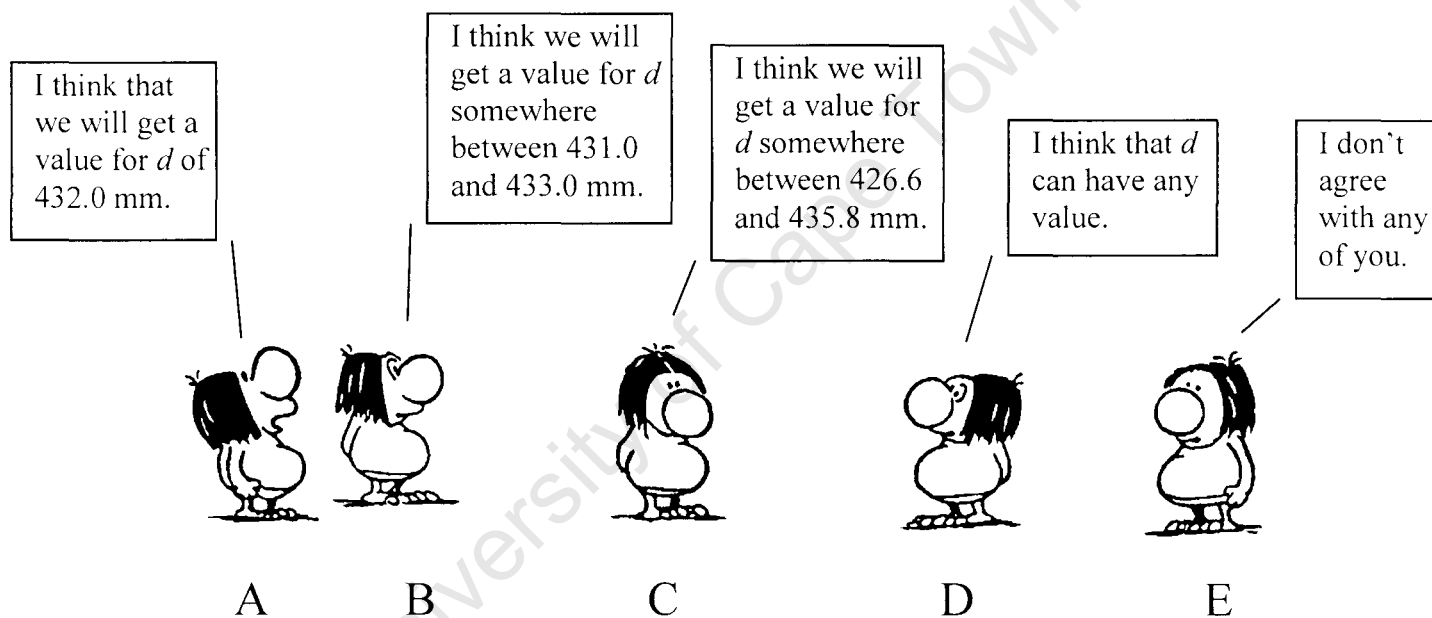
Q 5. (UA2/L)

The 5 values for d as shown below are the same as in the last question, i.e. obtained by the second group.

<u>Roll</u>	<u>d (mm)</u>
1	434.5
2	432.0
3	435.8
4	426.6
5	<u>431.1</u>

Average: **432.0**

The students now discuss what value they will get for d if they release the block again (for the sixth time) from point P .



With whom do you most closely agree? (Circle ONE):

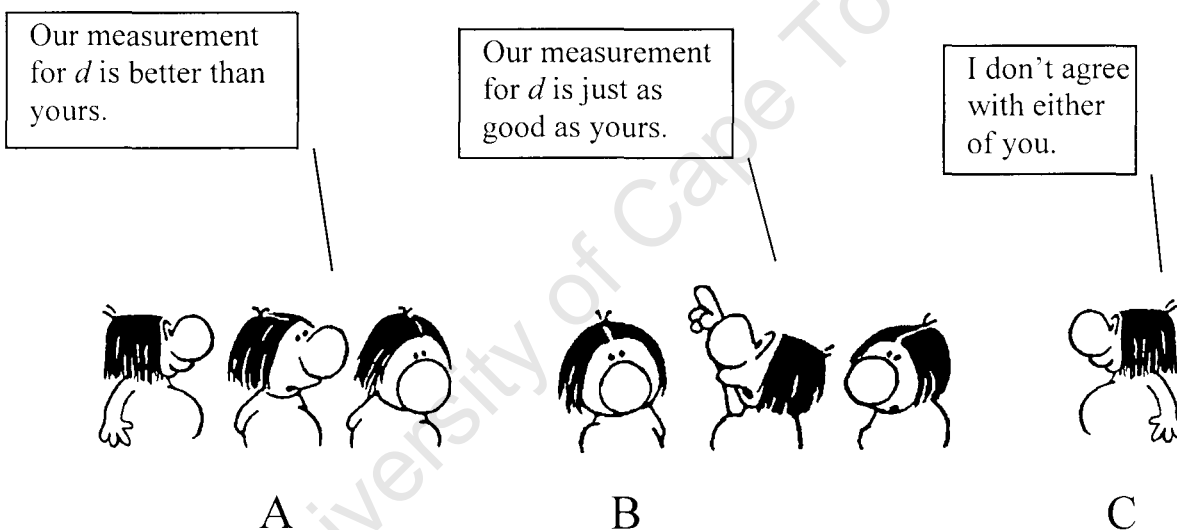
A	B	C	D	E
---	---	---	---	---

Explain your choice.

Q 6. (SMDS/L)

Two groups of students, both of whom had decided to release the block 5 times from point P , compare their measurements for d . Their values for the five releases are shown below, together with their averages.

Release	Group A d (mm)	Group B d (mm)
1	443.3	446.8
2	432.8	459.4
3	424.4	410.5
4	439.6	423.3
5	434.9	435.0
Average:	435.0	435.0



With whom do you most closely agree? (Circle ONE):

A	B	C
---	---	---

Explain your choice. Do not use the word “results or measurements” in your explanation: state clearly if you are referring to the “data values”, the “average”, etc.

Q 7. (DMSS/L)

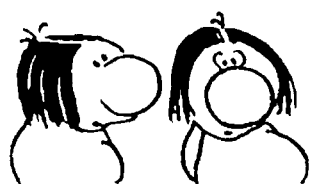
Two other groups of students compare their measurement of d . Their values for five releases are shown below, together with their averages.

<u>Release</u>	<u>Group A</u> <u>d (mm)</u>	<u>Group B</u> <u>d (mm)</u>
1	439.5	435.6
2	438.4	439.2
3	433.1	428.0
4	422.8	433.1
5	431.3	438.3
Average:	433.0	434.8

Our result for d agrees with yours.

No, our results do not agree.

I do not agree with either of you.



A

C



B



With which group do you most closely agree? (Circle ONE):

A

B

C

Explain your choice. Do not use the word “results or measurements” in your explanation: state clearly if you are referring to the “data values”, the “average”, etc.

Q 8. (ED1/L)

The lecturer now comes around with a special electronic meter which has a digital display and uses it to measure d . Here is what the electronic meter shows:



After recording the reading and the lecturer has gone, the following discussion takes place between the students.

Good, we now know that d is exactly 433.0 mm.

No, I think that we now know that d is approximately 433.0 mm.

I think that d is between 432.5 and 433.5 mm.

I think that d is between 431.0 and 432.0 mm.

I don't agree with any of you.

A B C D E

With whom do you most closely agree? (Circle ONE):

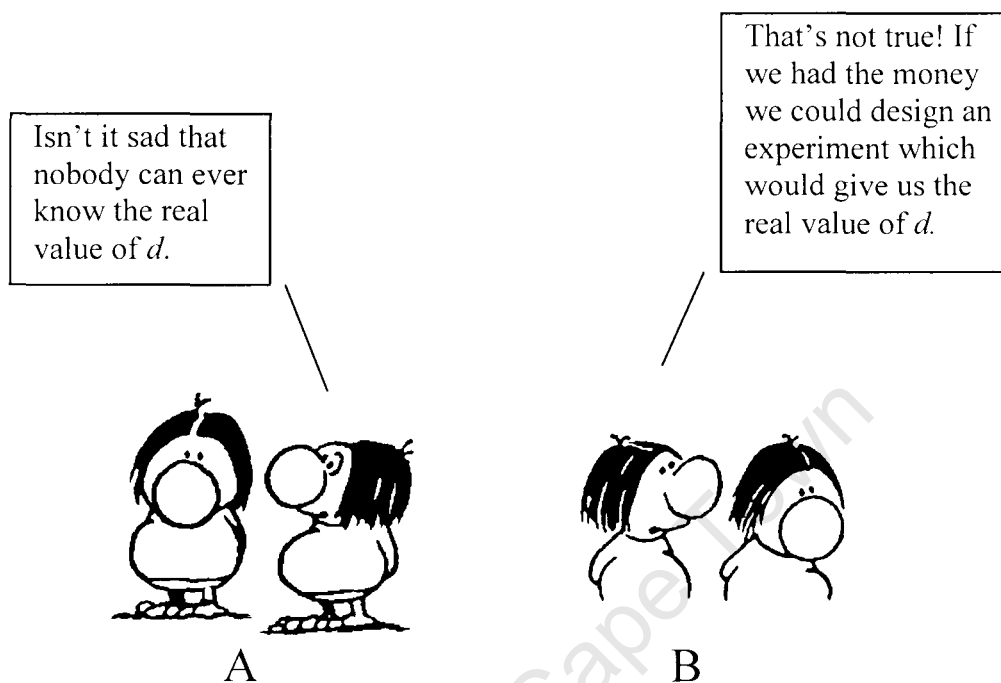
A	B	C	D	E
---	---	---	---	---

Explain your choice.



Q 9. (PX1/L)

The students continue to discuss the experiment.



With which group do you most closely agree? (Circle ONE):

A	B
---	---

Explain your choice.

Appendix IV

The coding schemes for the post-intervention probes

For each coding scheme, the table headings are:

Number : The number of student responses allocated to each category.

Code : The alphanumeric code for each category

P/S : The allocation of each category to either the point or set paradigm, if appropriate.

Category : A short description of each category.

University of Cape Town

Q 1. (PR1/L):

Number	Code	P/S	Category
0	N00	-	No response
1	U00	-	Not able to code response
	A		exactly 433.8 mm, because ...
0	A00	-	(no reason given)
0	A01	-	(not able to code reason given)
0	A12	P	the edge of the block is exactly in the middle.
0	A41	P	exact conditions for repeating will result in the exact same distance.
	B		approximately 433.8 mm, because ...
0	B00	-	(no reason given)
0	B01	-	(not able to code reason given)
0	B10	P	of the size of the spot.
0	B11	P	of the size/shape/movement of block.
0	B12	P	the edge of the block is between 433.5 and 434.0.
3	B13	P	closer markings/smaller intervals are needed, scale is too coarse.
2	B14	S	human judgement is required to estimate reading, error of parallax.
0	B15	P	a better measuring instrument is needed.
6	B21	S	measurement/the reading is not exact/perfect, so use 'approximate'.
0	B22	S	in Science/Physics/ Maths measurements are not exact, so use 'approximate'.
6	B23	P	the spot/reading is 'more or less' at 436 mm, so use 'approximate'.
0	B31	P	must repeat to find accurate answer.
0	B32	S	repeating will allow calculation of average.
0	B33	S	repeating will confirm range of measurements.
0	B40	S	the distance/position of spot is influenced by external factors.
1	B41	S	the distance/position of spot is influenced by experimental procedure/measurement process.
4	B50	S	it is the number that best represents the interval, or is the best estimate
5	B51	S	the measurement is approximate with respect to the best estimate and standard uncertainty.
3	B60	S	cannot know true value, use best estimate/approximation.

	C	between 433.0 and 434.0 mm because...
0	C00	- (no reason given)
0	C01	- (not able to code reason given)
0	C10	P of the size of the spot.
1	C11	P of the size/shape/movement of block.
6	C12	P the edge of the block is not exactly on a mark.
6	C13	P closer markings/smaller intervals are needed, scale is too coarse, need better calibration.
1	C14	S human judgement is required to estimate reading, error of parallax.
2	C21	S measurement is not exact, use interval to account for standard uncertainty.
0	C40	S external factors cause variations when repeating.
3	C50	S the measurement is in the interval.
4	C51	S the measurement is in the interval, therefore the distance is in the interval.
6	C52	S the measurement/reading is in the interval to account for standard uncertainty.
5	C53	S the true value of the distance is one number in the interval.
0	C54	S distance is best estimate plus standard uncertainty, in interval.
2	C61	S cannot know the true value, is in interval.
2	C62	S no exact/true value, interval accounts for uncertainties.
1	C63	S no exact/true value, best estimate is in interval.
	D	approximately 434.0 because...
0	D00	- (no reason given)
0	D01	- (not able to code reason given)
	E	I don't agree, because...
0	E00	- (no reason given)
0	E01	- (not able to code reason given)
1	E11	P of the size/shape/movement of block.
1	E14	S human judgement is required to estimate reading, error of parallax.
0	E30	U need to repeat measurements.
0	E31	S must repeat to get best estimate.
0	E32	S must repeat to calculate average.
0	E40	S must account for experimental uncertainties.
3	E50	S must calculate best estimate and standard uncertainty.
1	E52	S distance is in specified interval.
0	E60	U must round to the nearest 10.

Q 2. (PR2/L):

Number	Code	P/S	Category
0	N00	-	No response
0	U00	-	Not able to code response
	A		exactly 434.0 mm, because ...
0	A00	-	(no reason given)
0	A01	-	(not able to code reason given)
0	A11	P	the block has travelled exactly 434 mm.
5	A12	P	the edge of the block/the point Q is exactly on the 434 mm mark.
1	A13	P	the distance is easily read/seen/observed from scale on ruler.
0	A20	P	measurement is exact.
0	A41	P	exact conditions for repeating will result in the exact same distance.
	B		approximately 434.0 mm, because ...
0	B00	-	(no reason given)
0	B01	-	(not able to code reason given)
0	B11	P	of the size/shape/movement of block.
3	B12	P	the edge of the block/the point Q is not exactly on the 434 mm mark.
2	B13	P	closer markings/smaller intervals are needed, scale is too coarse.
5	B14	S	human judgement is required to estimate reading, error of parallax.
0	B15	P	a better measuring instrument is needed.
6	B20	S	measurement / the reading is best estimate, so use 'approximate'.
8	B21	S	measurement/ the reading is not exact/perfect, so use 'approximate'.
0	B22	S	Physics, (and/or Maths), requires exact numbers, but measurements are not exact, so use 'approximate'.
7	B23	P	the edge of the block is 'more or less'/approximately at 434.0 mm.
4	B24	S	measurement/the reading/distance is not exact/perfect, must calculate standard uncertainty.
0	B30	U	need to repeat measurements.
0	B31	P	repeating will give a recurring value.
0	B32	S	repeating will allow calculation of average.
0	B33	S	repeating will confirm range of measurements.
0	B34	S	repeating will not necessarily give same value
0	B40	S	the distance/position of block is influenced by external factors.
0	B41	S	the distance/position of block is influenced by experimental procedure/measurement process.
3	B50	S	434 mm is the best estimate, true value is in interval calculated taking standard uncertainty into account.
1	B60	S	true/exact value can't be found, use best estimate and associated uncertainty.
4	B61	S	true/exact value can never be known, so use 'approximate'.
2	B62	S	true /exact value can never be known since there is always uncertainty associated with experimental conditions/methods

	C	between 433.0 mm and 435.0 mm, because ...
0	C00	- (no reason given)
0	C01	- (not able to code reason given)
0	C11	P of the size/shape/movement of block.
1	C12	P the edge of the block/the point Q is not exactly on a mark.
2	C21	S reading/measurement is not exact/perfect, or is uncertain, so use interval.
1	C24	S measurement/the reading/distance is not exact/perfect, must calculate standard uncertainty.
0	C40	S the distance/position of block is influenced by external factors.
0	C41	S the distance/position of block is influenced by experimental procedure/measurement process.
2	C50	S best estimate is in given interval, must account for uncertainties.
2	C51	S reading is in given interval, therefore distance is in given interval.
4	C52	S distance is in another, specified interval.
0	C53	S interval accounts for uncertainties.
1	C61	S true/exact value can never be known, is in given interval
1	C62	S true /exact value can never be known since there is always uncertainty associated with experimental conditions/methods
	D	exactly 434 mm, because ...
0	D00	- (no reason given)
0	D01	- (not able to code reason given)
0	D11	P the block has travelled exactly 434 mm.
2	D12	P the edge of the block/the point Q is exactly on the 434 mm mark.
0	D13	P the distance is easily read/seen/observed from scale on ruler.
0	D20	P measurement is exact.
	E	I don't agree with any of you, because ...
0	E00	- (no reason given)
0	E01	- (not able to code reason given)
1	E14	S human judgement is required to estimate reading, error of parallax.
1	E20	S measurement is always uncertain, can't be sure.
1	E21	S there is no exact answer, (so use approximate)
0	E30	U repeats are necessary.
0	E35	S repeat to get good approximation/best estimate.
0	E40	S of external factors.
0	E41	S of experimental conditions.
4	E50	S reading is best estimate, distance is best estimate \pm standard uncertainty.
0	E51	S reading is in specified interval
0	E52	S distance is in specified interval
1	E60	S no true value, must calculate best estimate \pm standard uncertainty.
1	E61	S true/exact value can never be known, use approximate/best estimate

Q 3. (UR/L):

Number	Code	P/S	Category
0	N00	-	No response
0	U00	-	Not able to code response
14	10	S	average of the readings is final result for d
29	11	S	average of the readings is best estimate/most accurate value of d
2	12	P	average is the true/exact value of d
0	20	S	average of the readings, excluding the lowest reading, is final result for d
1	30	P	median reading is final result of d
14	40	S	average \pm standard uncertainty is final result for d
16	41	S	average \pm standard uncertainty, with a coverage probability, is final result for d

University of Cape Town

Q 4. (UA1/L):

Number Code P/S Category

0	N00	-	No response
0	U00	-	Not able to code response
	A		d is exactly 432 mm, because ...
0	A00	-	(no reason given)
0	A01	-	(not able to code reason given)
0	A10	P	average is the true distance
1	A30	P	average is exactly 432 mm.
1	A31	P	average is calculated using a formula, and is therefore correct.
0	A32	P	average remains the same after more repeats.
0	A41	P	if individual readings differ, then the average is the exact value.
	B		d is approximately 432 mm, because ...
0	B00	-	(no reason given)
0	B01	-	(not able to code reason given)
0	B10	S	average is closest to exact value/true distance.
6	B11	S	average is most likely value/best approximation of distance d .
1	B12	S	average is not necessarily the exact/actual/true distance.
1	B13	S	actual distance could be more or less than the average distance.
4	B16	S	average is best estimate, can never know true value.
1	B20	S	error must be accounted for.
2	B21	S	external factors must be accounted for.
4	B30	S	average is approximate, not exact.
1	B31	S	more repeats are necessary to find accurate average/answer.
2	B33	S	average will/may change after more repeats.
0	B41	S	individual readings/measurements are not identical, so there is no exact value.
0	B42	S	readings are close to 432 mm.
0	B44	S	no reading is 432 mm.
0	B50	P	of the shape/size/movement of the ball/block.
3	B70	S	of standard uncertainty.
1	B71	S	don't know exact interval in which d lies, must calculate standard uncertainty of average.
	C		d is between 431.5 mm and 432.5 mm, because ...
0	C00	-	(no reason given)
0	C01	-	(not able to code reason given)
1	C11	S	average is most likely value/best approximation of d , but not exact.
1	C13	S	actual distance is close to the average.
3	C14	S	actual distance is in interval.
2	C15	S	average is actual distance rounded off/ interval accounts for rounding off.
1	C16	S	can never be sure of the real d .
5	C20	S	allow/account for error/uncertainty.
3	C30	S	average is approximate/not exact/not accurate.
0	C33	S	average can/will change after more repeats.
7	C70	S	interval allows for standard uncertainty of average, (calculate to find best estimate and standard uncertainty)

	D	d is between 426 mm and 436 mm, because ...
0	D00	- (no reason given)
0	D01	- (not able to code reason given)
1	D11	S average in interval, so actual distance is also in interval.
3	D14	S actual distance in interval, average is best representation of interval/easy number to work with.
2	D20	S account for error/uncertainty.
0	D30	S average is approximate/not exact.
0	D31	S more repeats are necessary to find accurate average/answer.
1	D33	S average will change after more repeats.
0	D34	S average is in interval.
5	D40	S all readings are in that interval.
0	D41	S readings spread between 426 mm and 436 mm, so can't know exact distance.
0	D44	S no reading is 432 mm, so actual distance is described by the interval/range.
1	D70	S interval accounts for standard uncertainty.
2	D72	S interval found by calculating standard uncertainty of average, is in this interval.
	E	I don't agree with any of you, because ...
0	E00	- (no reason given)
0	E01	- (not able to code reason given)
0	E13	S actual distance is close to average.
0	E16	S average provides minimal value of actual distance.
1	E30	S average is approximate, not exact.
0	E31	S need more repeats to find accurate average/answer.
0	E32	P average is the same, whatever the readings.
2	E33	S average will change after more repeats.
0	E42	S readings are close to 432 mm.
0	E60	U average and actual d are unrelated.
5	E70	S d lies in uncertainty interval about average (best estimate), must calculate.
2	E71	S can't reach conclusion, need additional information about standard uncertainty.

Q 5. (UA2/L):

Number	Code	P/S	Category
0	N00	-	No response
0	U00	-	Not able to code response
	A		reading of 432 mm, because ...
0	A00	-	(no reason given)
0	A01	-	(not able to code reason given)
1	A30	P	average remains unchanged.
	B		reading between 431 mm and 432 mm, because ...
0	B00	-	(no reason given)
0	B01	-	(not able to code reason given)
1	B16	S	sixth reading will be close to fifth reading
2	B31	S	average between 431 mm and 432 mm.
1	B32	S	average remains close to 432 mm.
0	B40	S	reading must be close to average.
2	B41	S	many repeats stabilize average
	C		reading between 426 mm and 436 mm, because ...
0	C00	-	(no reason given)
0	C01	-	(not able to code reason given)
7	C10	S	all readings in range 426 mm-436 mm.
0	C11	S	readings have no set pattern.
0	C12	S	readings vary/change/are not the same.
1	C14	S	reading can have any value in range, can't predict exactly
3	C20	S	same experimental conditions, readings in same range.
1	C21	S	external factors affect readings.
1	C31	S	average is between 426 mm and 436 mm.
0	C41	S	more repeats will change average.
0	C50	S	434 mm occurs twice, reading likely to be close.
1	C61	S	next reading in interval given by standard uncertainty about average.
1	C70	S	interval allows for standard uncertainty of average, (calculate to find best estimate and standard uncertainty)
	D		reading can have any value, because ...
0	D00	-	(no reason given)
0	D01	-	(not able to code reason given)
1	D11	S	readings have no set pattern, so can't predict.
2	D12	S	readings vary/change/are not all the same, so can't predict.
7	D13	S	reading can have any value, can't be predicted.
2	D14	S	reading can have any value in specified/unspecified bigger ranger.
0	D15	S	reading can have any value, close to previous values.
5	D20	S	experimental factors affect readings.
18	D21	S	external factors affect readings.
0	D40	S	reading will be close to average.
5	D41	S	reading will change average, so can't predict.

	E	I don't agree with any of you, because ...
0	E00	- (no reason given)
0	E01	- (not able to code reason given)
1	E10	S can't predict sixth reading
1	E12	S readings vary/change/not the same , so can't predict.
2	E14	S reading can have any value in specified bigger range.
1	E15	S reading can have any value close to previous values, can't predict exactly.
2	E20	S experimental conditions affect readings.
1	E21	S external factors affect readings.
0	E30	P average remains unchanged.
1	E40	S readings must be close to average.
2	E41	S reading will change average, can't predict.
1	E42	S reading will be greater than average.
0	E60	S can't know true value of d
2	E61	S can't know true value of d, need to calculate best estimate and standard uncertainty.

University of Cape Town

Q 6. (SMDS/L):

Number	Code	P/S	Category
1	N00	-	No response
0	U00	-	Not able to code response
A A's results are better, because ...			
0	A00	-	(no reason given)
0	A01	-	(not able to code reason given)
7	A10	S	they have a smaller range/spread
1	A11	S	they have a smaller range/spread because of outside factors
4	A12	S	they have smaller range/spread because fewer mistakes were made
2	A13	S	they have smaller range/spread , therefore a more accurate/reliable average
0	A14	S	they have smaller range/spread, therefore are closer to true value
0	A15	S	they have smaller range/spread because group A was more skilful
2	A16	S	they have smaller range/spread, and therefore a smaller uncertainty
0	A20	S	there is less deviance from the average
0	A21	S	there is less deviance from the average because of outside factors
0	A22	S	less deviance from the average because of fewer mistakes made
0	A25	S	less deviance from the average because group A was more skilful
0	A40	U	you usually get the results so close together
0	A50	P	their average (435 mm) is also one of the measurements
4	A62	S	A's data values are closer to each other
4	A63	S	A's data values are more accurate/ consistent
3	A64	S	A's data values are closer to their average
B B's results are just as good as A's, because ...			
0	B00	-	(no reason given)
0	B01	-	(not able to code reason given)
1	B10	S	they got more or less the same data values
1	B11	S	the spread/range of data values overlap
1	B12	S	sets of data values are equally good - same no. of readings obtained under same experimental conditions
7	B20	P	they have the same average
1	B21	P	they have the same average although different outside factors caused deviation
0	B22	P	they have the same average although mistakes caused deviation
3	B23	P	they have the same average, and the spread is not important
2	B24	S	they have the same average, and the same standard uncertainty
1	B26	P	they have the same average, deviation not important as expected
9	B27	P	they have the same average, individual data values not important
0	B29	P	they have the same average and same number of readings
0	B30	P	they have the same average, although A got 435 mm on their last measurement.
0	B60	U	there is no exact answer to an experiment like this
1	B61	P	there is no exact/true answer, the average is the best estimate and is the same
1	B64	P	different data values but same average
0	B65	P	the accuracy of individual readings is not under consideration, the average is important
1	B70	P	it is a natural outcome of the same experiment, the spread is not important)
1	B81	S	the intervals give by the standard uncertainties about the averages overlap

	C		I don't agree with either of you because ...
0	C00	-	(no reason given)
0	C01	-	(not able to code reason given)
2	C10	S	the data values are far apart/differ
1	C12	S	different spread/range of data values, depend on experimental factors
0	C40	U	you usually get the results so close together
2	C60	P	don't know true value, so can't compare to averages.
2	C66	S	data values inconsistent - affected by external factors
10	C80	S	need standard uncertainties to compare
0	C90	U	B better than A
1	C91	S	B better than A, bigger range but same average

University of Cape Town

Q 7. (DMSS/L):

Number	Code	P/S	Category
0	N00	-	No response
0	U00	-	Not able to code response
	A		Our results <u>agree</u> with yours, because ...
0	A00	-	(no reason given)
0	A01	-	(not able to code reason)
2	A10	P	the readings/measurements for both sets are more or less the same
0	A12	S	the readings/measurements for both sets have the same spread
6	A13	S	the readings/measurements have an overlapping spread
1	A20	P	the averages are more or less the same
0	A21	P	the averages are more or less the same, difference due to external factors
1	A22	P	the averages are more or less the same, difference due to experimental errors
0	A24	P	the averages are more or less the same, both close to true value
0	A26	P	the averages are more or less the same as there will always be deviation
14	A30	S	the uncertainties of the averages may overlap
4	A31	S	the averages are more or less the same with similar ranges/spreads
0	A40	P	three out of five (the majority) of readings are the same
0	A50	P	if you round off the averages, then they are identical
	B		No, your results <u>do not agree</u> with ours, because ...
0	B00	-	(no reason given)
0	B01	-	(not able to code response)
0	B10	P	the readings are not the same.
1	B12	S	the spreads of both sets are different
6	B20	P	the averages are different
1	B21	P	the averages are different due to different conditions/external factors
1	B22	P	the averages are different due to experimental errors
0	B24	P	the averages different – uncertain about where the true value lies
0	B25	P	the averages are different, absolute accuracy/identical results required to agree
0	B26	P	the averages are too different even though deviation is taken into consideration
3	B30	S	the averages are too far apart for the uncertainties to overlap
1	B31	P	average is different and all individual readings are not the same
3	B32	S	the spread differs between the two
1	B40	P	both groups got some different measurements
0	B50	P	if you round off the averages, then they are very different
0	B60	P	an average is only true if the average value also appears as one of the measurements
0	B80	P	group B is more accurate than group A

	C		I don't agree with either of you
0	C00	-	(no reason given)
1	C01	-	(not able to code response)
8	C20	S	can't compare averages, need standard uncertainties.
15	C30	S	intervals may/may not overlap - need standard uncertainties to compare.
3	C40	P	groups have some common data values
1	C60	P	don't know true value, so can't compare to averages.
1	C70	S	different experimental technique - need standard uncertainties to account, then compare.
2	C71	S	different experimental conditions - need standard uncertainties to account, then compare.

University of Cape Town

Q 8. (ED1/L):

Number	Code	P/S	Category
0	N00	-	No response
0	U00	-	Not able to code response
	A		exactly 433.0 mm, because ...
0	A00	-	(no reason given)
0	A01	-	(not able to code reason given)
0	A11	P	the ball has travelled exactly 433.0 mm.
3	A12	P	the reading is 433.0 mm.
1	A15	P	electronic meter gives exact/accurate answers.
0	A41	P	exact conditions for repeating will result in the exact same distance.
	B		approximately 433.0 mm, because ...
0	B00	-	(no reason given)
0	B01	-	(not able to code reason given)
0	B10	P	of the size of the spot.
0	B11	P	of the size/shape/movement of block.
0	B12	P	the reading is approximate.
0	B13	P	meter not sensitive enough, need better calibration.
0	B14	S	human judgement is required to estimate reading, error of parallax.
0	B15	P	a better measuring instrument is needed.
6	B21	S	measurement/the reading is not exact/perfect, so use 'approximate'.
0	B22	S	in Science/Physics/Maths measurements are not exact, so use 'approximate'.
1	B24	S	measurement close to true value, so use approximate.
0	B30	U	need to repeat measurements.
0	B31	P	repeating to confirm, get recurring value.
0	B32	S	repeating to account for scatter by calculation of average.
1	B34	S	there are variations when repeating.
1	B40	S	variations caused by external factors.
2	B60	S	can never know true value, always uncertainty.
6	B62	S	can never know true value, uncertainty associated with internal calibration of meter.
6	B63	S	can never know true value, uncertainty associated with sensitivity of meter.

	C	between 432.5 mm and 433.5 mm, because ...
0	C00	- (no reason given)
1	C01	- (not able to code reason given)
3	C13	P need better calibration, more sensitive.
0	C15	P need better measuring device, meter inaccurate.
1	C16	S uncertainty associated with reading off meter
7	C20	S reading is not exact/reading is best estimate, use interval to account for uncertainties.
2	C21	S meter gives reading which is best estimate, can't be sure
1	C33	S repeats will confirm range, readings will fall in interval.
2	C40	S of external factors
0	C41	S of experimental method/measuring process.
7	C50	S must calculate standard uncertainty to get best approximation of d.
2	C51	S must calculate standard uncertainty associated with digital scale to get best approximation of d.
0	C60	S can never know true value, always uncertainty.
5	C62	S can never know true value, uncertainty associated with internal calibration of meter.
7	C63	S can never know true value, uncertainty associated with sensitivity of meter.
	D	between 431 mm and 432 mm, because ...
0	D00	- (no reason given)
1	D01	- (not able to code reason given)
	E	I don't agree with any of you, because ...
0	E00	- (no reason given)
0	E01	- (not able to code reason given)
0	E15	S of mistakes of the measuring device.
0	E20	S measurement is always uncertain, can't be sure.
1	E21	S meter gives reading which is best estimate, can't be sure
0	E30	U repeats are necessary.
0	E31	P repeat to confirm, find recurring value.
0	E32	S repeat to account for scatter and calculate average.
0	E34	S repeat to account for external factors and approximate.
0	E40	S of external factors.
4	E50	S must calculate standard uncertainty to get best approximation of d.
1	E51	S must calculate standard uncertainty associated with digital scale to get best approximation of d.
0	E60	S can never know true value.
2	E61	S can never know true value, calculate most likely interval.
1	E62	S can never know true value, uncertainty associated with internal calibration of meter.
1	E63	S can never know true value, uncertainty associated with sensitivity of meter.

Q 9. (PX1/L):

Number	Code	P/S	Category
0	N00	-	No response
0	U00	-	Not able to code response
	A		Can never know the real value of d, because ...
0	A00	-	(no reason given)
0	A01	-	(not able to code reason)
2	A10	S	all measurements are approximate.
9	A20	S	true value can't be known.
16	A21	S	true value can't be known, only approximate value/best estimate.
6	A22	S	true value can't be known, only most likely interval.
27	A30	S	there are always uncertainties, can't be reduced to zero.
1	A31	S	there are always uncertainties due to experimental conditions.
12	A32	S	there are always uncertainties due to external factors.
1	A33	S	there are always uncertainties due to human judgement.
35	A40	S	can never have a good/sensitive enough scale/measuring device.
	B		If we had money we could design an experiment to give the real value of d, because...
0	B00	-	(no reason given)
1	B01	-	(not able to code response)
6	B10	P	create an instrument good/sensitive enough to measure true/exact/real value.
1	B11	P	create an instrument good/sensitive enough so that uncertainty can be ignored.